

Strategic Evaluations: World Bank Responses to the Founding of China's AIIB *

Jing Qian, James Raymond Vreeland, Jianzhi Jason Zhao

September 2025

Abstract

China's new Asian Infrastructure Investment Bank (AIIB) represents a potential challenge to the World Bank: AIIB founding countries appear to have turned away from some World Bank projects in favor of working instead with the AIIB. We examine the World Bank's response to this competition, focusing on its efforts to retain the participation of the AIIB founding members. Drawing on the accommodation dilemma framework of Dellmuth and Walter (2025), we explore World Bank *ex post* project evaluations. Positive evaluations offer recipients certain benefits that may appease the AIIB founders—but are subtle enough to minimize contagion risk, lest the World Bank's accommodation of AIIB founders provoke similar actions from other countries. Our analysis reveals that World Bank projects for AIIB founding members have systematically received *higher* evaluations since the establishment of the Chinese-led institution. The study provides insights into the strategic use of nuanced policy instruments and the dynamics of accommodating countries in the context of contested multilateralism.

*Jing Qian, Assistant Professor Faculty Fellow of Political Science, New York University, Shanghai; Email: jingqian@nyu.edu. James Raymond Vreeland, Professor, Princeton University; Email: james.raymond.vreeland@gmail.com. Jianzhi (Jason) Zhao, Associate Professor, University of Exeter; Email: j.zhao@exeter.ac.uk.

1 Introduction

The Asian Infrastructure Investment Bank (AIIB) opened its doors on January 2016 with 57 founding members. Led by China and positioning itself as a more inclusive and efficient alternative, the AIIB quickly emerged as a competitor to the World Bank in financing critical infrastructure projects (Qian et al. 2023). The establishment of this new development bank thus signals a bold challenge to the current financial order—a shift in the approach that countries dissatisfied with the Bretton Woods institutions are seeking to renegotiate the terms of global development finance (Lipsky 2017; Pratt 2021). For many, the AIIB offers a fresh option to bypass the World Bank’s lengthy processes and stringent conditionality, setting the stage for a new era of contested multilateralism.

The creation of the AIIB has thus forced the World Bank to grapple with the “accommodation dilemma” (Dellmuth and Walter 2024; Walter 2023, 2021b; Jurado et al. 2022a; Walter and Plotke-Scherly 2025). When cooperation is threatened, defectors may be won back through accommodation—but only at the risk of incentivizing others, who would like similar accommodation, to defect as well (see, for example, Walter 2021a; Hobolt et al. 2022; Malet and Walter 2024). In this case of the WB and the AIIB, the cooperation gains at stake for the World Bank are potentially high. Among the 57 founders are 27 World Bank borrowers, including large economies such as India and Indonesia; combined, the founders represent 82 percent of World Bank borrower GDP.¹ This is not a marginal group, and in the first three years of its existence, these AIIB founders turned away from World Bank project loans, reducing their annual number of new World Bank infrastructure projects by more than 20 percent (Qian et al. 2023, 219). Such a trend, were it to continue, could disrupt the operations of the World Bank, which depend on interest paid on project loans. Thus accommodation is warranted to win back the AIIB founders.

Contagion, however, is a concern. Overt accommodation of the AIIB founders could damage the reputation of the World Bank—and encourage new countries to defect. Coun-

1. Here, we define “borrower” as any country that has taken an IBRD loan or IDA credit in the quinquennium before the founding of the AIIB in 2016 (i.e., 2011) through the present. To calculate World Bank borrower GDP, we use the World Bank’s World Development Indicators data in current prices from 2023 (there is typically a two-year gap for data reporting). For Ethiopia and Yemen there is no data reported, and we use IMF data. Neither source has data for Eritrea, and so we use the UNTAD 2023 report.

tries might shift their borrowing for infrastructure needs to the AIIB or even seek, someday, to establish other banks in hope of extracting the accommodation benefits enjoyed by the AIIB founders.

We thus suggest that the key for the World Bank is to rely on a strategy of accommodation that is both *meaningful* to recipients but also *unnoticed* by other governments. In this way, the World Bank can induce the return of AIIB founders, who are drifting away, without incentivizing new countries to follow the same strategy as the founders. Otherwise, the World Bank will have to offer similar inducements to all countries.

What options does the World Bank have to accommodate AIIB founders without provoking contagion? Larger project loans with weakened oversight, for example, would be valuable to AIIB founders, but would carry a high risk of contagion. By contrast, some sort of lip service to general improvements in project delivery would not risk contagion, but could be viewed as cheap talk by AIIB founders. Project evaluations, however, may offer a strategic tool that the World Bank can use to thread the needle.

Project evaluations matter to governments. They play a meaningful role in shaping a country's ongoing borrowing relationship with the World Bank. They can also influence the level of domestic politics credit a government can claim for the benefits of a project to the country.

At the same time, however, evaluations fly below the radar. They are typically perceived simply as technical instruments for assessing project outcomes. In contrast to international press coverage of project loans, which is commonplace, ex-post evaluations of project performance are not typically discussed in global news. Favorable evaluations thus offer a form of accommodation of AIIB founders without drawing unwanted attention from other member states. They represent a nuanced, low-cost mechanism for managing the dual imperatives of retaining the AIIB founders and maintaining institutional credibility.

We offer both qualitative and quantitative evidence supporting our argument. Qualitative evidence from three interviews with World Bank economists suggests the plausibility of our argument. Analysis of quantitative data further reveals that the projects of AIIB founding members—compared to other projects—receive higher ratings after the AIIB's es-

tablishment in 2016. We also find that these evaluations are less likely to be downgraded in subsequent reviews by the World Bank’s Independent Evaluation Group (IEG). This set of results suggest that the World Bank’s evaluation practices may reflect a calculated strategy of accommodation in response to institutional competition.

Our study provides empirical evidence on an adaptive strategy of the World Bank in an era of contested multilateralism (Morse and Keohane 2014). The findings build on literature showing how IOs respond to the emergence of alternative institutions (e.g., Lipsky 2017). Moreover, our work provides an empirical application of the accommodation dilemma theoretical framework in the narrow but important context international organizations (IOs) (Dellmuth and Walter 2025; Walter 2023). In particular, we present a case where the level of contagion risk is endogenous to the form that accommodation takes. The research further contributes to the broader literature on IO responses to non-cooperation, showing how institutional actors balance flexibility and credibility in maintaining their influence within an evolving global order.

2 The World Bank and the Founding of the AIIB

The World Bank, founded at the close of World War II, has long been the preeminent multilateral development finance institution. With 189 member countries and a workforce exceeding 12,000 full-time staff, its mission is to eradicate extreme poverty and promote sustainable development (World Bank 2024). The institution also served as the template for the major regional multilateral development bank (MDB)s: the African, Asian, and Inter-American Development Banks, and the European Bank for Reconstruction and Development (AfDB, AsDB, IDB, and EBRD) (Heldt and Schmidtke 2019). Collectively, these institutions have played a dominant role in international development cooperation (Ahluwalia et al. 2016; Humphrey 2022, 74).

Over the decades, the World Bank has faced mounting criticism. Scholars and policymakers have highlighted issues such as unsustainable development policies (Park 2007; Weaver 2008), prolonged project approval processes (Humphrey 2015), and insufficient financing for critical infrastructure needs (Kellerman 2019). More controversially, the World

Bank's reliance on conditional lending, often tied to neoliberal economic reforms aligned with the Washington Consensus, has provoked resistance among borrowing countries (Clark and Dolan 2021). These critiques underscore a growing dissatisfaction among developing countries.

A core grievance is the imbalance of influence within the Bank. Developing countries lack the voting power necessary to effect change (Lipsky 2017; Pratt 2021). Western powers, led by the United States, dominate the Bank's governance structure. While the voice of the Global South is an issue at all of the legacy MDBs, underrepresentation is most pronounced at the World Bank (Humphrey 2022, 115).

Against this backdrop of dissatisfaction, the AIIB was established with 57 founding members spanning five continents. Its establishment represents "the most dramatic shift in the multilateral development landscape for decades" (145).² The institution is marked by its distinctive scale and ambition, the absence of the United States and Japan among its members, and, crucially, its leadership under China (Qian et al. 2025, 1076).

Part of the AIIB's appeal lies in its promises of inclusivity and responsiveness. Research shows that countries underrepresented in the existing international financial system were more likely to join as founding members (Wang 2018). At the AIIB, they enjoy increased voting power relative to their share at the World Bank (Kim and Lee 2020).

Operationally, the AIIB adopts a streamlined approach, neatly aligned with the immediate needs of many developing countries. Branded as a "lean" and "clean" institution, the AIIB emphasizes efficiency and focuses on infrastructure financing with few conditions attached (Zhao et al. 2019). This contrasts with the World Bank's lengthy approval procedures and extensive social and environmental standards (Park 2007).

Sometimes portrayed as "China's World Bank," the AIIB poses a challenge to the World Bank.³ The threat to the World Bank is not just theoretical. AIIB founding members began to shift their engagement away from the World Bank soon after its operations commenced in 2016, specifically pursuing fewer infrastructure loans (Qian et al. 2023).

2. See also Rodrigues Vieira (2018).

3. See, for example, "China Creates a World Bank of Its Own, and the U.S. Balks," *New York Times* (December 4, 2015), "The AIIB: China's World Bank," *Bloomberg* (April 16, 2015) and "China's World Bank alternative points to multilateral future without US," *S&P Global* (July 17, 2019).

These first-mover countries defied the public opposition of the United States when they helped found the AIIB.⁴ Their dissatisfaction with the World Bank and interest in working with China’s new institution were sincere. Years of ignoring complaints have left the World Bank at risk of continuing to lose influence in this important set of countries.

3 How the AIIB Provokes an Accommodation Dilemma

The accommodation dilemma refers to trade-offs that arise when responding to non-cooperation (Walter 2020, 2021a, 2021b; Jurado et al. 2022b). Non-cooperation in international relations occurs when states or groups of states seek to challenge the status quo (Dellmuth and Walter 2025; Walter and Plotke-Scherly 2024). Accommodation of non-cooperative actors involves changing costs and benefits—shifting payoffs favorably towards them in hope of maintaining their support. As noted, such accommodation can trigger contagion, where new actors become non-cooperative in order to extract similar payoffs. Actors privileged by the status quo must weigh the benefits of preserving cooperation—the gains at stake—against the reputational risks of accommodating non-cooperative behavior (Dellmuth and Walter 2025, 10). If the cooperation gains at risk are low—and especially if the contagion risk is high—actors may pursue only low levels of accommodation or no accommodation at all.

The establishment of the AIIB represents a specific form of non-cooperation—contested multilateralism—characterized by the founding of an organization that challenges the existing institutional order: a “competitive regime” (Morse and Keohane 2014, 392). To be clear, it reflects a form of *constructive* non-cooperation. This contrasts with outright defection, where governments may withdraw from participation in an international organization or a treaty altogether (Dellmuth and Walter 2025, 2).⁵ With the founding of the AIIB, states

4. See, for example, Yang and Van Gorp (2019, 615–616). This opposition was evident in October 2014 during the AIIB’s initial stages, just before the signing of the Memorandum of Understanding (MOU) for establishing the China-led MDB. The United States actively lobbied its allies against joining the AIIB (“[Big nations snub Beijing bank launch after US lobbying](#),” *Financial Times*, October 22, 2014). The tension between Washington and Beijing intensified so much that Jin Liqun, who would eventually become the first AIIB president, explicitly requested the US Ambassador to China to moderate US opposition (“[US Opposing China’s Answer to World Bank](#),” *New York Times*, October 9, 2014). Despite Jin’s plea, US opposition persisted. As one observer put it, the United States “forced allies and friendly countries across the Far East to make a fatal choice between the US and China” (“[US Risks Epic Blunder by Treating China as an Economic Enemy](#),” *The Telegraph*, March 25, 2015).

5. See also Cooley and Nexon (2020), Schmidt (2025), and Von Borzyskowski and Vabulas (2025).

have remained members of the World Bank. They, arguably, seek to renegotiate terms, address perceived injustices, and reform the existing system from the outside (Dellmuth and Walter 2025, 6). By creating an alternative MDB, AIIB founders aimed to remedy perceived inefficiencies and inequities in the status quo without entirely abandoning multilateralism. Many observers—as well as the Asian Development Bank and the AIIB itself—pointed to a funding gap of more than \$700 billion for infrastructure needs in Asia (Desai and Vreeland 2015). While the explicit goal of the AIIB is to complement the World Bank, the institution implicitly offers an alternative more in line with the preferences of many developing countries. This constructive approach allows dissatisfied actors to challenge the dominance of existing IOs while maintaining a commitment to cooperative problem-solving.

Still, competitive regime creation, even in this constructive form, creates winners and losers (Morse and Keohane 2014). For the World Bank, the AIIB threatens to erode its authority and status as the leading MDB (Katada 2016; Gray 2018, 2024; Schmidt 2024). Accordingly, the World Bank has engaged in “intense discussions at the senior level” about competition from the AIIB (Zaccaria 2024, 171). The founders are too important to lose, but the World Bank must avoid incentivizing contagion, whereby other countries drift toward the AIIB to reap similar accommodation benefits of accommodation. In short, the World Bank must find a way to maintain its relevance in the face of non-cooperation.

Accommodation of the AIIB founders helps the World Bank maintain lending relationships with a key group of developing countries, who are pivotal for its operational sustainability as well as its legitimacy as the world’s leading development bank. Yet, accommodation might create the expectation that challenging the World Bank is a winning strategy, thereby encouraging future non-cooperative actions (see Paradise 2019; Xiao 2015).

Dellmuth and Walter (2025, 10) explain that navigating through the accommodation dilemma hinges on two key considerations: (1) preserving cooperation gains and (2) managing contagion risk. For the World Bank, the long-run cooperation gains are pivotal to maintain its leading role in economic development. We thus predict accommodation.⁶ We

6. To be clear, choosing to punish the AIIB founders at this juncture would only further push them away from the World Bank and towards the AIIB (Daßler et al. 2024, see). Losing these key borrowers to the AIIB risks further erosion of the World Bank’s influence and relevance. The majority of the top borrowers at the World Bank are also AIIB members. For the IBRD, the top five borrowing countries in the 2024 fiscal year are

further suggest, however, that the risk of contagion is endogenous to the form of accommodation offered to defecting actors. Specifically, we proffer that positive evaluations grant benefits to AIIB founders that may not be noticed by other countries.

The level of contagion risk depends on the accommodation strategy. What options does the World Bank have? One lever the World Bank could pull is increasing the size of loans offered to AIIB founding members to maintain their cooperation. But this move would surely attract attention, risking contagion. Providing larger loans would be easily noticed and other borrowers would see temporary defection as a way to obtain such a desirable benefit.

On the other hand, the World Bank could accommodate AIIB founders by revising its lending conditions—and high standards—to align more closely with the AIIB’s simplified, adaptive practices, which have fewer contingencies and strings attached (Apolinario Junior and Jukemura 2022). Evidence does suggest that the World Bank has previously reduced conditionality in response to competition from China’s bilateral foreign aid (Hernandez 2017). However, using this approach in a targeted manner to accommodate AIIB founding countries carries a high risk of contagion—as well as reputational damage. If the World Bank relaxed conditionality or lowered standards exclusively for AIIB founders, other countries would notice. It would incentivize other countries to engage in similar forms of contestation, and thus increase the likelihood of further challenges to the World Bank’s authority. And if lenient conditions and standards were applied universally, it would undermine the World Bank’s reputation for maintaining rigorous lending practices and development standards. Such a shift could erode the Bank’s credibility and its long-standing role as a normative leader in global development finance.

Evaluations provide a pragmatic middle ground. As we explain below, they enable the World Bank to signal flexibility and responsiveness without compromising its core prin-

Ukraine, Türkiye, Indonesia, India, and the Philippines. For the IDA, they are Ethiopia, Bangladesh, Nigeria, Pakistan, and Tanzania. Among them, only Ukraine, Nigeria, and Tanzania are not yet AIIB members. See (World Bank 2024). That said, the United States publicly opposed countries’ joining the AIIB in the run-up to its founding in 2016 (Freeman 2019; Yang and Van Gorp 2019), and might be reluctant to accommodate AIIB as founders. A recent study by Qian et al. (2025) shows that the United States punished its allies who joined the AIIB founding members by casting non-affirmative votes on World Bank project proposals. However, even here, the punishment was gestural because the projects were approved despite the US vote—a slap on the wrist, rather than a harsh punishment.

ples or governance structures. Favorable evaluations convey the Bank's commitment to fostering successful partnerships, and strengthen borrower loyalty, which may mitigate competitive pressures posed by the AIIB. By granting more positive evaluations to AIIB founders, the Bank can strategically cultivate stronger future relationships with these key borrowers, safeguarding cooperative gains that might otherwise be at risk. Offering more positive evaluations allows the Bank to accommodate AIIB founders without openly signaling institutional weakness or compromising its foundational standards.

And while evaluations may be touted in domestic media, they are rarely covered by international press. They are unlikely to garner the attention of other member states. Indeed, one would be hardpressed to even detect bias of this nature on a case-by-case basis. Tailoring evaluation outcomes thus helps manage contagion risks, as favorable evaluations are perceived as context-specific rather than universal concessions to non-cooperative behavior. This approach reduces the likelihood of other borrowers interpreting such actions as precedents for leniency or adjustments in response to external pressures.

Finally, unlike adjustments to lending volumes or conditionality, evaluations require no structural reforms or substantial financial commitments, making them a cost-effective tool for addressing the accommodation dilemma.

Among the various accommodation strategies available to the World Bank, adapting evaluation outcomes stands out as uniquely suited to addressing the challenges posed by the AIIB. Evaluations offer a subtle yet impactful mechanism to navigate the complex dynamics of cooperation and competition, balancing the need to retain borrower loyalty with the imperative to uphold institutional credibility.

4 Project Evaluations as Accommodation

4.1 Brief background on World Bank Project Evaluations

In 1973, the World Bank launched the Operations Evaluation Department (OED), the first fully independent evaluation unit among IOs—with a direct reporting line to the Board of Executive Directors (Jankauskas and Eckhard 2023, 25). Since the 1980s, the World Bank has

required an evaluation report for all projects. The commitment to independent oversight was further consolidated in 2005 through the establishment of the IEG, which unified the evaluation functions across the various facets of the World Bank Group, institutionalizing a more cohesive and rigorous approach to project and program assessment. Since then, the World Bank has established harmonized evaluation criteria across its operational and evaluation units (IEG 2015, 1).⁷

For the sample used in our empirical analyses below, the Bank mandates a self-evaluation called an Implementation Completion and Results Report (ICR). The ICR process is initiated by the task team leader overseeing the project. Once finalized, the ICR is approved by the country director, submitted to the Board, and then made publicly available (World Bank 2011).

It is important to note that there may be systematic differences across the ICR scores of projects financed through the Bank's two main lending windows—the International Bank for Reconstruction and Development (IBRD) and the International Development Association (IDA). While both financing arms follow the same core evaluation procedures, they operate under distinct incentive structures. IBRD projects are typically implemented in middle-income countries and carry market-based terms, while IDA provides concessional financing to low-income countries and allocates resources based on performance. As a result, projects funded under these two arms may face different pressures in the evaluation process, and the interpretation of ratings may vary accordingly.

For IDA countries, a central institutional mechanism is the Country Performance Rating (CPR)⁸, which plays a critical role in allocating concessional resources. Updated annually, the CPR is derived from the Country Policy and Institutional Assessment (CPIA) and project-level outcomes. These ratings not only guide the distribution of scarce IDA resources but also inform broader strategic decisions at the World Bank. Project evaluations in IDA countries thus carry heightened financial and reputational consequences, creating stronger incentives for borrower governments and Bank staff to avoid negative assessments. In con-

7. Over the years, the specific types and names of these evaluation reports have undergone several changes to adapt to evolving institutional priorities and methodologies. For a detailed overview of the evolution of IEG's project-level evaluations, see IEG (2015, 2–3). See also Kilby and Michaelowa (2019).

8. See <https://ida.worldbank.org/en/financing/resource-management/ida-country-performance-ratings>

trast, IBRD projects do not face such direct performance-based allocation mechanisms. This institutional asymmetry motivates our disaggregated analysis by lending type, as the stakes of accommodation may be more acute in IDA settings.

Finally, it is important to recognize that each World Bank evaluation includes distinct sub-ratings for “borrower performance” and “Bank performance,” which reflect different dimensions of accountability. The borrower rating assesses how well the recipient government and implementing agencies fulfilled their responsibilities, while the bank rating evaluates the World Bank’s own performance in project design, supervision, and support. Our core analysis centers on the overall project outcome ratings, which reflect the World Bank’s assessment of development effectiveness. We also disaggregate the analysis by examining sub-ratings of borrower and bank performance. This allows us to assess whether the observed accommodation effect is concentrated in assessments of the recipient government’s role or the Bank’s own implementation record—providing insight into whether favorable evaluations for AIIB founders reflect shifts in attribution rather than systematic inflation of the Bank’s self-assessment.

The IEG also *independently* validates each ICR to assess both the performance of the project and the quality of the World Bank’s self-evaluation (World Bank 2011). The Implementation Completion and Results Report Review (ICRR) is conducted for all ICRs. The ICRR is a desk-based review of the ratings presented in the ICR (IEG 2014).

In addition to these desk-based reviews, the IEG further conducts in-depth, field-based evaluations for approximately 20 to 25 percent of projects, resulting in Project Performance Assessment Report (PPARs). The selection of projects for PPARs is nonrandom and guided by several considerations, including the IEG’s work program priorities, the potential for disputed ratings, and the opportunity for future learning and innovation.

PPAR evaluations provide a more detailed and comprehensive analysis of project performance and outcomes.⁹ PPAR evaluations offer a more rigorous and independent assessment of project outcomes than ICRRs. As desk-based validations, ICRRs review the plausibility of the self-assessed ratings and narrative provided in the ICR. This review only relies exist-

9. See <https://ieg.worldbankgroup.org/methodology/PPAR>. See also Kilby and Michaelowa (2019).

ing on project documentation and a final interview with the task team leader. In contrast, PPARs are full evaluations. They involve new evidence collection through country missions, site visits, and interviews with a wide range of stakeholders—government officials, civil society, and beneficiaries. PPARs use a mixed-methods approach that combines literature review, portfolio analysis, and fieldwork. Ratings from PPARs override those from the ICRR. Because of their depth and field-based methodology, PPARs are generally considered more credible.¹⁰ They are often used as building blocks for broader country and corporate evaluation studies (IEG 2015, 2). These differences help explain why PPAR-based results may better reveal political or strategic influences on project evaluation.

In our empirical analyses, we distinguish between the two evaluation types. All projects receive an ICRR rating, which serves as the default IEG score. For the subset with PPARs, we follow IEG practice and use the PPAR rating to supersede the ICRR rating. This design allows us to compare outcomes across distinct review processes and assess whether the World Bank’s accommodation is more visible under closer scrutiny.

4.2 The Relevance of Evaluations

World Bank project evaluations matter to governments. These assessments serve as inputs for the World Bank’s broader strategic planning.¹¹ According to an IEG survey, the World Bank Executive Board frequently uses evaluation findings to assess ongoing engagement with member states (IEG 2019, 36). As we learned from our interviews, poor performance ratings, especially downgrades imposed by the IEG, can influence the loans a country receives.¹²

Beyond their importance for government interactions with the World Bank, evaluations may also matter to governments at the level of domestic politics. While not typically covered internationally, evaluations may be reported in the media domestically, where they matter. A study of the World Bank-funded KALAHI-CIDSS project in the Philippines, for example, demonstrates that politicians can leverage foreign aid projects for electoral advan-

10. See <https://ieg.worldbankgroup.org/methodology/PPAR>

11. See Ravallion (2016), Kilby and Michaelowa (2019), World Bank (2011), IEG (2014), and World Bank (2014). See also “World Bank country assistance strategies,” *The Bretton Woods Project* (April 16, 2010).

12. Interview A.

tages through undeserved credit claiming (Cruz and Schneider 2017). Of course, a necessary condition to claim credit for success is evidence of a positive evaluation. The KALAHICIDSS project received a “satisfactory” evaluation in both its self-evaluation and the IEG validation.¹³ Philippine government agencies publicized these positive evaluations to their domestic public, framing them as evidence of effective governance and development outcomes. The Department of Social Welfare and Development published “World Bank satisfied with welfare projects,” and the Department of Agriculture published “World Bank rates DA-PRPD satisfactory.”¹⁴

Conversely, unsatisfactory ratings can attract negative coverage and public criticism in domestic media, as evidenced in India and Pakistan. The “moderately unsatisfactory” progress rating for Kerala’s solid waste management project led to scrutiny of local governance in India, while the closure of a locust control project in Pakistan due to poor performance drew similar public criticism.¹⁵

Borrowing countries accordingly perceive evaluation outcomes as markers of their success in using World Bank resources. They matter both for governments’ relationship with the World Bank Executive Board and for their domestic politics.

4.3 AIIB Founders and World Bank Project Evaluations

Although evaluations are often regarded as impartial assessments, they are not insulated from political pressures. They can be used as instruments in institutional power struggles (Wildavsky 1972, 516). This is particularly evident when evaluation outcomes influence the future allocation of resources, potentially altering power dynamics among stakeholders (Wergin 1976). In such cases, political actors may exert pressure on evaluators to “misrepresent findings” or adjust methodologies to produce favorable results (Pleger et al. 2017, 316). These dynamics underscore the complexity of evaluation processes and the potential for

13. See <https://projects.worldbank.org/en/projects-operations/project-detail/P077012>.

14. See <https://www.dswd.gov.ph/world-bank-satisfied-with-welfare-projects/>, Department of Social Welfare and Development, Republic of the Philippines (July 10, 2013); and <http://prdp.da.gov.ph/world-bank-rates-da-prpd-satisfactory/>, Department of Agriculture, Republic of the Philippines (February 15, 2018).

15. “World Bank rates Kerala solid waste management project’s progress moderately unsatisfactory”, *The Indian Express* (August 17, 2024); “World Bank closes \$200m locust project”, *DAWN* (January 22, 2024).

bias to reflect broader institutional and political imperatives.

Evaluation biases are a well-documented phenomenon specifically within IOs. Political factors influence the design, implementation, and interpretation of evaluation outcomes (Jankauskas and Eckhard 2023). Similar patterns of bias have been observed in foreign aid agencies, where evaluations can be shaped by external pressures and organizational objectives (Michaelowa and Michaelowa 2011; Wood and Howes 2025). A significant driver of such biases is the dual role that evaluations play: beyond assessing project performance, they serve as tools of self-legitimation. Evaluations are frequently employed as marketing devices to “prove” the success of an organization’s initiatives to the general public (Michaelowa and Borrman 2006).

That said, many studies analyzing World Bank project evaluations treat the ratings as proxies for performance, while acknowledging their limitations.¹⁶ Still, there is growing evidence that these evaluations are subject to political distortions and potential biases. For instance, Malik and Stone (2018) show that projects involving Fortune 500 multinational corporations as contractors are more likely to receive inflated ratings. Similarly, Kilby and Michaelowa (2019) find that countries holding nonpermanent seats on the United Nations Security Council at the time of evaluation are more likely to receive higher IEG ratings in PPAR evaluations. These findings suggest that geopolitics shape not only the allocation of World Bank loans but also the evaluations of funded projects.¹⁷

We contend that evaluations are not merely objective measures of project effectiveness; they are also shaped by the strategic imperatives and broader goals of the institution. There are compelling incentives for the World Bank to assign higher evaluation ratings to AIIB founders in particular. As noted earlier, ICR ratings are prepared by the operations staff and are often led by task team leaders responsible for supervising the projects. These leaders have an incentive to maximize project lending and disbursement volumes, which align with their performance metrics and organizational priorities (Weaver 2008). Faced with a declin-

16. See, among others, Honig (2019), Heinzel et al. (2023), Honig et al. (2023), Heinzel and Liese (2021), Heinzel (2022), Watkins (2022), Winters (2019), Shin et al. (2017), Bulman et al. (2017), Girod and Tobin (2016), Kilby (2015), Dreher et al. (2013), and Denizer et al. (2013).

17. On geopolitics and World Bank lending, see, e.g., Dreher et al. (2009), Fleck and Kilby (2006), and Kersting and Kilby (2016).

ing borrowing trend from AIIB founders, staff in World Bank country offices are motivated to “win back more projects” from these key borrowers (Qian et al. 2023, 230).

Moreover, AIIB founders possess enhanced bargaining power, which they can leverage during the ICR preparation process. Borrowing countries provide critical inputs and comments that shape the evaluation, giving them indirect influence over the final ratings. This dynamic further supports the expectation that AIIB founders would receive higher evaluations in the ICR process.

Hypothesis 1: AIIB founders are more likely to receive higher ICR ratings for World Bank projects evaluated after 2016.

While the operations staff may have incentives to inflate ICR evaluations, these tendencies can be moderated by the validation and auditing process conducted by the IEG (Malik and Stone 2018). The IEG often revises original ICR ratings, with downgrades being more frequent than upgrades (Malik and Stone 2018; Kilby and Michaelowa 2019). As an independent body reporting directly to the Board of Executive Directors, the IEG’s structure provides a degree of insulation from operational biases.

These downgrades matter to upper management and the Executive Board. As one of our interviews revealed, “if projects from a particular country are repeatedly downgraded in IEG reviews or in the Project Performance Assessment Reports (PPARs), the department head—or even the Board—may begin to question whether that country should continue to receive such a large share of future allocations.”¹⁸

Importantly, however, the independence of the IEG does not fully preclude the influence of geopolitics. Kilby and Michaelowa (2019) demonstrate that IEG evaluations favor countries holding nonpermanent seats on the United Nations Security Council, for example. Similar geopolitical dynamics may apply to AIIB founders. If so, these countries would be less likely to experience rating downgrades from the IEG. The IEG may strategically adjust its evaluations to accommodate AIIB founders while maintaining its broader institutional objectives.

18. Interview A.

Hypothesis 2: AIIB founders are less likely to receive rating downgrades from the IEG for World Bank projects validated after 2016.

5 Interview Evidence

We interviewed three World Bank economists. Interview subject A is a lead economist at a World Bank Country Office, with experience working in multiple countries, including the Washington, DC headquarters. Interview subject B is a senior economist at a World Bank Country Office, also with experience working at the Washington, DC headquarters. Interview subject C is a former specialist at the Washington, DC headquarters, who now works for another MDB.

Interestingly, the interviews with two higher ranking economists with in-country experience working for the World Bank (Interviews A and B) offer evidence supportive of our hypothesis. The interview with the specialist with experience in Washington only, who has since left the World Bank (Interview C), offers starkly less support. We suspect that there is good reason for this pattern. First, the results from interview C reveal that political influence over World Bank evaluations may fly below the radar. Second, the influence of AIIB founding members may be recent, emerging after our interview subject left the World Bank.

Accordingly, when asked whether political considerations influence ICR results, Interview C answered, “I am not aware of any cases where bilateral relationships have affected ICR results. We assess these results based on the project itself and through regular consultations with our clients... we base our ratings on experience and facts, which are typically agreed upon with our clients. While they may not always agree with our ratings—and often feel the ratings are not high enough—they tend to request a higher evaluation.”

By contrast, Interview A answered, “the challenge lies in the fact that almost all indicators in the ICR are not entirely objective... Political considerations can also play a role, especially when a project is rated as ‘unsatisfactory’—which tends to be more contentious. That’s when politics or power dynamics must come into play. If the borrower country is more powerful or holds greater bargaining power, I can imagine they are more likely to push back or even the board or vice presidency will support them and thus less likely to end

up with a negative rating.”

Similarly, Interview B answered, “The entire project cycle requires maintaining a good relationship with our borrowers. Their domestic political situation can influence project implementation and outcomes. International politics also affects lending—it first shapes the Country Partnership Framework and the overall portfolio allocated to each country, which in turn can influence ICR results.”

When asked more directly about our hypotheses about AIIB founding members, Interview C was again skeptical, suggesting that their influence might, at best, be “limited.” Yet, the subject also stated, “I don’t have much personal experience with this, but it is true that our borrowers have more options these days. I’m not very familiar with how this competition is affecting our operations.”

By contrast, Interview A replied, “I believe it aligns with what I mentioned earlier. I can imagine that these countries are less likely to receive problematic ratings.” Similarly, Interview B answered, “Now that borrowers can access funding more easily from institutions like the AIIB, our lending staff face increased pressure to lend more. As a result, they may be inclined to relax certain standards to justify additional lending.”

Interview B further offered a rhetorical question whose implication conforms closely with the logic of accommodation, “Imagine a situation where two similar projects are implemented by the same borrower—one with the World Bank that receives a ‘Moderately Satisfactory’ rating, and the other with the AIIB that receives a ‘Highly Satisfactory’ rating. Which institution do you think the government official would prefer to work with next time?”

We observe a similar pattern from our questions about rating downgrades from the IEG. Interview C was dismissive: “To be honest, as one of the lending staff, I’m not actually concerned that much about the results from IEG. I cannot remember any IEG rating put us in an awkward position or caused embarrassment to our department or people I know. Nor I think any IEG rating is affecting our lending or interaction with our borrowers.”

By contrast, Interview A suggested, “the [IEG] rating is something that the Bank—especially at the top level—cares about. I also believe client countries care about the results.” The

subject elaborated, “Political factors could indeed be involved, particularly when the IEG downgrades projects too severely, too frequently, or when outcome indicators are rated as unsatisfactory.” Interview B explained, “it’s possible that IEG staff may be reluctant to present themselves as obstructing World Bank lending—especially in an increasingly competitive development finance landscape. Some in the operations department do feel that IEG is holding them back. This perception might influence IEG to avoid downgrading projects in countries where there is greater competition.”

We conclude from our interviews that our hypotheses are plausible. Interviews A and B indicate that AIIB members may receive special treatment; interview C confirms that, if so, the benefits are subtle enough to go unnoticed. We conclude from our qualitative evidence that the political influence the AIIB members has likely influenced the evaluation process in some cases. We turn to quantitative evidence to better understand whether this practice adheres to systematic patterns.

6 Research Design

6.1 Data on World Bank Project Evaluations

This study draws on project evaluation ratings from the World Bank API, which provides data for both self-evaluated ICR reports and IEG assessments.¹⁹ Using a single, consistent source aligns variable definitions and coding across ICR and IEG evaluations, enabling direct comparison in our empirical analysis. Multiple versions of the IEG Project Performance Database are available on the World Bank’s platform, differing mainly in project coverage and the range of variables provided. To maintain full comparability with the ICR data, we rely on the current API-based version

The IEG dataset contains evaluation outcomes from both desk-based ICRR and field-based PPAR for selected projects. When multiple evaluations are available for the same project, we follow scholarly convention and [IEG \(2015\)](#) guidance by using the most recent

19. See <https://search.worldbank.org/api/v3/projects>. Our process involved feeding the API with a comprehensive list of project IDs from the universe of World Bank projects. See <https://projects.worldbank.org/en/projects-operations/projects-list>.

evaluation. We also explore differences between ICRR and PPAR outcomes. We refer to these data as our IEG dataset.²⁰

Note that the ICR dataset is key both for examining potential biases in the World Bank’s self-evaluation process and to serve as a baseline for analyzing the validations in the IEG dataset. Despite its importance, prior research has largely overlooked ICR data, primarily due to challenges in accessibility.²¹ This paper is the first to conduct a direct analysis of self-evaluated ICR ratings and to compare them to IEG validations, providing a novel contribution to the literature on multilateral development bank accountability and evaluation practices.

For both ICR ratings and IEG evaluations, this study focuses on the overall project performance rating, defined by the IEG as “the extent to which the operation’s major relevant objectives were achieved, or are expected to be achieved, efficiently” (IEG 2015, 14). The original performance ratings are assigned on a six-point scale, ranging from 1 (very unsatisfactory) to 6 (very satisfactory). To facilitate analysis, we use both the original ratings and a binary transformation, where a value of 1 indicates a project rating of moderately satisfactory, satisfactory, or highly satisfactory, and 0 otherwise. We also examine whether patterns differ between borrower performance and bank performance ratings, which serve as sub-indexes of project performance.

6.2 Coding AIIB Founding Membership

The AIIB’s founding members are explicitly listed in Schedule A of its Articles of Agreement. To capture the interaction between AIIB founding membership and the timing of evaluations, we construct the variable *AIIB founder* \times *Post-2016*. This variable is coded as 1 for projects evaluated in 2016 or later that involve AIIB founding members, reflecting the establishment of the AIIB in 2016. For all other projects—those evaluated before 2016 or

20. The IEG dataset is widely used in scholarly research. For studies of project performance, see Winters (2019), Shin et al. (2017), Bulman et al. (2017), Kilby (2015), and Denizer et al. (2013). For studies of World Bank effectiveness, see Heinzel et al. (2023) and Heinzel (2022). For studies of borrowing countries, see Watkins (2022), Heinzel and Liese (2021), and Girod and Tobin (2016).

21. An exception is Malik and Stone (2018), who compare ICR evaluations with performance ratings coded by research assistants. Kilby and Michaelowa (2019) report having used ICR ratings, but their data appears to be sourced from the IEG dataset, suggesting that they (understandably) relied on ICR Review (ICRR) evaluations by the IEG—not the original ICR ratings.

involving non-AIIB members—this variable is coded as 0.²²

6.3 Control Variables

Our analysis includes several control variables at both the project and country-year levels to account for factors that may influence World Bank project evaluation outcomes.

Because prior research highlights the importance of project-level “micro” variables in explaining variation in project outcomes that could influence evaluations (Denizer et al. 2013; Bulman et al. 2017), we control for total project cost (log-transformed) and sector fixed effects.²³ We do not control for project duration, as it is potentially endogenous—problematic projects often take longer to complete, making duration partly an outcome of the same processes affecting evaluations. Nonetheless, the results remain substantively unchanged when project duration is included as an additional control.

Additionally, for the analysis of IEG ratings, we include a control for the type of evaluation report—whether the evaluation is based on an ICRR or a PPAR. This distinction is important, as PPARs typically involve field-based assessments and reflect a more rigorous evaluation process compared to desk-based ICRRs. We also conduct subsample analyses by evaluation type to explore whether the association between AIIB founding membership and evaluation outcomes differs systematically between ICRR- and PPAR-based assessments.

Several country-year-level “macro” variables may influence World Bank project evaluations. We control for GDP per capita and total population, both logged, to account for the level of economic development and country size.²⁴ Building on the work of Girod and Tobin (2016)—who emphasize the significance of government income sources and fiscal constraints in explaining compliance with World Bank project agreements—we include net foreign direct investment (FDI) inflow as a percentage of GDP and total natural resource rent as a percentage of GDP.²⁵ These variables reflect dependency on external and non-tax fi-

22. For robustness, we also replicate the analysis using the year of formal membership for AIIB founders, as well as using non-founders as the comparison group.

23. Sector information is collected from the World Bank API. For each project, we identify the major sector accounting for the largest share of the World Bank’s appraisal cost.

24. Data sourced from the World Development Indicators, World Bank, <https://datacatalog.worldbank.org/dataset/world-development-indicators>.

25. Data sourced from the World Bank, <https://data.worldbank.org>.

nancial flows, which may influence project implementation and evaluation outcomes. To account for the quality of democratic institutions and the prevalence of corruption, both of which may influence borrowing patterns, we include the polyarchy index and public corruption indicator from the Varieties of Democracy (V-Dem) Project (Coppedge et al. 2024; Pemstein et al. 2024). All country-level variables are measured one year prior to the project's approval date, ensuring that they reflect conditions at the time of project preparation and avoid potential bias from changes during implementation.

Lastly, Kilby and Michaelowa (2019) highlight a potential link between geopolitics and World Bank project evaluations, particularly for borrowing countries that are geopolitically significant at the time of evaluation. To account for the alignment of geopolitical interests, we use the ideal point distance between the United States and the borrowing country, as formulated by Bailey et al. (2017). We also include an indicator variable that equals 1 if the borrowing country holds a nonpermanent seat on the United Nations Security Council (UNSC) during the evaluation year (Dreher et al. 2009). Both variables are measured at the year of evaluation to reflect geopolitical relevance at the time the project is assessed.

6.4 Sample

The dataset covers World Bank project evaluations for all IBRD and IDA projects evaluated between 2007 and 2020.²⁶ We exclude projects that: (1) are designated as regional rather than country-specific; (2) lack information on evaluation outcomes or dates; or (3) have IEG evaluation dates preceding their respective ICR evaluation dates.

The resulting ICR sample includes 3,192 projects across 138 countries, 26 of which are AIIB founding members. Of these, 3,021 projects have both ICR and IEG outcome ratings, also spanning 138 countries, including 26 AIIB founding members.

26. For ICR data, only 27 projects were evaluated in 2006, with the earliest dated September 28; these were excluded from the analysis.

6.5 Specification

To examine the relationship between AIIB founding membership and World Bank project evaluations, we estimate the following ordinary least squares regression model with multi-way clustered standard errors:

$$\begin{aligned} \text{Evaluation}_{ict_a, t_e} = & \beta \text{AIIB Founder}_c \times \text{Post-2016}_{t_e} \\ & + \gamma X_{ic, t_a-1} + \delta Z_{ic, t_e} + \eta W_i + \alpha_c + \theta_{t_e} + \phi_{\text{sector}(i)} + \epsilon_{ict_a, t_e} \end{aligned}$$

In this model, $\text{Evaluation}_{ict_a, t_e}$ represents the evaluation rating for project i implemented in country c , approved in year t_a , and evaluated in year t_e . The primary variable of interest, $\text{AIIB Founder}_c \times \text{Post-2016}_{t_e}$, is a binary indicator that equals 1 for AIIB founding members whose projects were evaluated in 2016 or later, and 0 otherwise.

The matrices X_{ic, t_a} and Z_{ic, t_e} include country-year level control variables measured at the year prior to project approval ($t_a - 1$) and evaluation (t_e), respectively; W_i contains project-level control variables. The model includes country fixed effects (α_c), year fixed effects for the year of evaluation (θ_{t_e}), and sector fixed effects ($\phi_{\text{sector}(i)}$). The term ϵ_{ict_a, t_e} denotes the error term, clustered at both the country and evaluation-year levels (Cameron et al. 2011).

7 Results

The findings provide evidence supporting our hypotheses: AIIB founding members are more likely to receive higher project evaluations for their World Bank projects evaluated after 2016. These elevated ratings tend to persist through subsequent IEG validation, with patterns suggesting that the IEG does not systematically counteract the initial advantage—and in some instances, may reinforce it. Notably, original staff ratings for projects involving AIIB founding members appear, in some cases, less likely to be downgraded upon review. That said, there are nuances to these overall results.

7.1 Baseline Results: ICR Evaluation

Table 1 presents the results using data from ICR evaluation ratings. In columns 1–3, the dependent variable is a binary indicator coded as 1 if the project evaluation falls within the “satisfactory” range. Column 1 includes only AIIB founding membership status and fixed effects for country and year as explanatory variables. Column 2 adds project-level controls, including the log of total project cost and sector fixed effects. Column 3 reports the full specification, incorporating both country-level and project-level covariates. The estimated coefficients for $AIIB\ Founder \times Post - 2016$ are consistently positive across these models, as expected; however only the estimate in column 2 reaches conventional levels of statistical significance, providing partial support for hypothesis 1.

Columns 4–6 shift the analysis to the original 6-point scale ratings from project evaluations. Here, the coefficients for $AIIB\ Founder \times Post - 2016$ are statistically significant in all models, suggesting that AIIB founding members tend to receive higher ratings when evaluations are assessed on a more granular scale. This finding confirms the expectation of hypothesis 1.

Table 1: AIIB Founders and ICR Evaluation

	ICR Evaluation (Binary)			ICR Evaluation (Scale)		
	(1)	(2)	(3)	(4)	(5)	(6)
AIIB Founder x Post-2016	0.041 (0.024)	0.041* (0.019)	0.041 (0.033)	0.111* (0.061)	0.108* (0.051)	0.119* (0.060)
Project cost (log)		0.032*** (0.007)	0.033*** (0.007)		0.107*** (0.020)	0.112*** (0.019)
GDP per capita (log)			0.024 (0.071)			0.095 (0.117)
Population (log)			0.131 (0.151)			0.370 (0.413)
FDI inflow (% GDP)			-0.001 (0.001)			-0.002 (0.002)
Natural resource rent (% GDP)			0.000 (0.003)			0.004 (0.006)
Corruption			-0.192 (0.117)			-0.343 (0.260)
Democracy			-0.177 (0.131)			-0.566** (0.259)
UNGA voting (Ideal Point distance from US)			-0.017 (0.048)			-0.044 (0.100)
Temporary UNSC member			0.047 (0.027)			0.143*** (0.047)
Observations	3192	3192	3049	3192	3192	3049
Countries	138	138	122	138	138	122
AIIB Founders	26	26	26	26	26	26
R2	0.096	0.110	0.106	0.142	0.164	0.166
R2 Adj.	0.052	0.062	0.058	0.100	0.119	0.121
Country Fixed Effects	✓	✓	✓	✓	✓	✓
Evaluation Year Fixed Effects	✓	✓	✓	✓	✓	✓
Sector Fixed Effects		✓	✓		✓	✓

Notes: Results from ordinary least squares regression. Robust standard errors clustered at country and evaluation-year levels reported in parentheses.

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

Taken together, the results in Table 1 indicate that while AIIB founding members are more likely to receive higher ICR ratings, the observed effect is primarily *within* the respective ranges of “satisfactory” and “unsatisfactory” ratings, rather than *across* the threshold marking the distinction of “satisfactory.” This implies that World Bank lending staff adopt a nuanced strategy when accommodating projects for AIIB founding members. This finding is also consistent with what we found in Interview A, “Political considerations can also play a role, especially when a project is rated as ‘unsatisfactory’—which tends to be more

contentious.”

7.2 Robustness of the ICR Findings

As a first robustness check, we adjust the timing of treatment to reflect each country’s formal ratification date of AIIB membership, rather than using 2016 as a uniform start point. While most AIIB founding members completed ratification by 2016, there are exceptions—Brazil, for instance, ratified in late 2020, and South Africa has not yet done so despite being listed in the Articles of Agreement. In our main analysis, we treat all countries listed as founders as treated post-2016 to reflect their political alignment at the institution’s inception. Nonetheless, using country-specific formal membership dates as the treatment onset (Table A.2) produces substantively similar results, suggesting that the observed evaluation patterns are not sensitive to this coding decision.

Second, we examine the set of non-founding members who joined the AIIB after its launch. Unlike founders, these countries were not critical to the institution’s formation and thus lacked the strategic leverage central to the accommodation dilemma. Consistent with our theoretical expectations, we find no evidence of elevated ICR ratings for non-founders (Table A.3). This contrast underscores the logic of targeted accommodation: only actors whose defection could jeopardize the cooperation gains or trigger broader reputational contagion appear to receive systematically more favorable evaluations. Exploring the mechanisms behind the relationship between AIIB founding membership and higher ICR project evaluations, we consider two potential mechanisms below.

7.3 Explore the Mechanism: IDA vs IBRD Loans

First, we examine whether the observed accommodation effect is concentrated among IDA projects. As discussed above, the IDA allocates funding using a distinctive performance-based formula—the Country Performance Rating (CPR)—which is unique within the World Bank system and directly determines country-level funding envelopes.²⁷ The CPR aggregates a range of indicators, with the greatest weight placed on 16 policy-related criteria

27. See <https://ida.worldbank.org/en/financing/resource-management/ida-country-performance-ratings>

grouped into four thematic clusters (World Bank 2021). Importantly, about 8% of the CPR is consistently allocated to the Portfolio Performance Rating (PPR), which reflects the share of problem projects in a country's IDA portfolio.²⁸ Although the underlying PPR data are not publicly available, IDA documents suggest that outcome ratings of closed IDA projects play a central role in constructing the PPR score (World Bank 2018, 85).

This institutional design provides a plausible channel through which evaluation ratings might be strategically adjusted. If effects are driven primarily by IDA projects, it would suggest that the accommodation operates through a mechanism that indirectly improves CPR scores, thereby enhancing future allocation prospects. This would align with our broader argument: while the Bank publicly adheres to a rules-based regime to uphold its credibility and manage reputational risk, it may retain informal means—such as internal evaluation ratings—to accommodate influential members, particularly those that joined as AIIB founders.

Table 2 presents subsample results for IDA and non-IDA projects. The dependent variable is the ICR's original 6-point evaluation score. In columns (1) to (3), focused on IDA projects, the estimated coefficients for AIIB Founder \times Post - 2016 are consistently positive and statistically significant. In contrast, for non-IDA projects (columns 4–6), the coefficients are small in magnitude and not statistically distinguishable from zero.

These results suggest that the observed accommodation is indeed concentrated among IDA projects—precisely where evaluation scores are most likely to influence resource allocation decisions.

28. See <https://thedocs.worldbank.org/en/doc/79cd8ebb4261ba33ba5f2968f783657c-0410012020/original/ida19-replenishment-report-annex-2.pdf>

Table 2: By IDA Funding: AIIB Founders and ICR Evaluation

	ICR Evaluation (Scale)					
	IDA Projects			Not-IDA Projects		
	(1)	(2)	(3)	(4)	(5)	(6)
AIIB Founder x Post-2016	0.194** (0.078)	0.195** (0.075)	0.228** (0.082)	0.041 (0.128)	0.060 (0.128)	-0.054 (0.132)
Project cost (log)		0.051*** (0.009)	0.056** (0.024)		0.159*** (0.035)	0.158*** (0.037)
GDP per capita (log)			0.052 (0.181)			0.263** (0.091)
Population (log)			0.348 (0.455)			0.852 (0.958)
FDI inflow (% GDP)			-0.003 (0.002)			-0.006 (0.012)
Natural resource rent (% GDP)			0.006 (0.007)			-0.003 (0.012)
Corruption			-0.479* (0.259)			0.584 (0.332)
Democracy			-0.382 (0.267)			-0.694* (0.324)
UNGA voting (Ideal Point distance from US)			-0.025 (0.118)			-0.047 (0.312)
Temporary UNSC member			0.024 (0.092)			0.251** (0.087)
Observations	1965	1965	1842	1226	1226	1206
Countries	94	94	94	66	66	66
AIIB Founders	22	22	22	17	17	17
R2	0.156	0.171	0.174	0.176	0.233	0.236
R2 Adj.	0.108	0.118	0.120	0.115	0.168	0.168
Country Fixed Effects	✓	✓	✓	✓	✓	✓
Evaluation Year Fixed Effects	✓	✓	✓	✓	✓	✓
Sector Fixed Effects		✓	✓		✓	✓

Notes: Results from ordinary least squares regression. Robust standard errors clustered at country and evaluation-year levels reported in parentheses.

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

7.4 Exploring the Mechanism: Borrower versus Bank Performance Ratings

Next, we explore whether the observed ratings advantage for AIIB founding members is more pronounced in evaluations of borrower performance or bank performance, using the corresponding sub-ratings from ICR evaluations. This distinction is substantively important. The borrower performance rating captures the extent to which the recipient govern-

ment and implementing agencies ensured high-quality preparation, implementation, and compliance with project agreements. In contrast, the bank performance rating reflects the World Bank's own role in ensuring quality at entry and providing effective supervision and support throughout the project lifecycle (IEG 2015).

This disaggregation is motivated by the idea that if the Bank is strategically adjusting evaluations to retain or reward influential borrowers, it is more likely to do so by adjusting how outcomes are attributed—downplaying borrower shortcomings or emphasizing their strengths—rather than altering assessments of its own staff. Such behavior would be consistent with the broader political logic of evaluation accommodation, particularly when borrowers have access to alternative sources of financing and therefore greater leverage in their relationship with the Bank (Watkins 2022; Girod and Tobin 2016).

Table 3 reports the results. Across models (1)–(3), which use borrower performance as the dependent variable, the estimated coefficients for the interaction term AIIB Founder \times Post-2016 are consistently positive and statistically significant in the first two specifications. In contrast, models (4)–(6), where bank performance is the outcome, yield smaller and statistically insignificant estimates.

These results suggest that the higher overall ICR ratings documented earlier are primarily driven by more favorable evaluations of borrower performance, rather than improvements attributed to the World Bank's own role. While we cannot directly observe the internal deliberations of evaluation teams, this pattern is consistent with the notion that the Bank may be directing “credit” toward AIIB founders for project outcomes, potentially as a signal of continued partnership quality. This is especially noteworthy given that borrower performance assessments are intended to reflect government capacity and compliance—dimensions likely to be of heightened relevance in the context of geopolitical competition and development finance fragmentation.

Table 3: Borrower & Bank Performance: AIIB Founders and ICR Evaluation

	ICR Evaluation (Scale)					
	Borrower Performance			Bank Performance		
	(1)	(2)	(3)	(4)	(5)	(6)
AIIB Founder x Post-2016	0.190** (0.079)	0.166* (0.082)	0.150 (0.106)	0.077 (0.060)	0.068 (0.059)	0.026 (0.063)
Project cost (log)		0.118*** (0.017)	0.114*** (0.017)		0.113*** (0.024)	0.109*** (0.021)
GDP per capita (log)			0.401** (0.142)			0.450*** (0.145)
Population (log)			0.749* (0.387)			1.109** (0.377)
FDI inflow (% GDP)			-0.008* (0.004)			-0.006*** (0.002)
Natural resource rent (% GDP)			0.001 (0.006)			0.001 (0.005)
Corruption			-0.398* (0.214)			-0.052 (0.275)
Democracy			-0.362** (0.158)			-0.086 (0.242)
UNGA voting (Ideal Point distance from US)			-0.012 (0.163)			-0.170 (0.104)
Temporary UNSC member			0.141* (0.069)			0.088* (0.049)
Observations	2665	2665	2546	3190	3190	3047
Countries	133	133	118	138	138	122
AIIB Founders	25	25	25	26	26	26
R2	0.189	0.218	0.230	0.154	0.184	0.196
R2 Adj.	0.142	0.169	0.181	0.111	0.140	0.152
Country Fixed Effects	✓	✓	✓	✓	✓	✓
Evaluation Year Fixed Effects	✓	✓	✓	✓	✓	✓
Sector Fixed Effects	✓	✓	✓	✓	✓	✓

Notes: Results from ordinary least squares regression. Robust standard errors clustered at country and evaluation-year levels reported in parentheses.

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

7.5 Evidence on IEG Downgrades

We next turn to the important evaluations conducted by the IEG. Recall that the IEG independently reassesses project outcomes using either desk-based ICRRs or in-depth, field-based PPARs. Table 4 examines whether these IEG reviews assign lower ratings than the original ICR evaluations.

We define downgrades using two measures: a binary indicator for shifts across the sat-

isfactory/unsatisfactory threshold and a more granular dummy variable capturing any decrease in the evaluation score.²⁹ Consistent with the general pattern reported by Kilby and Michaelowa (2019) and Malik and Stone (2018), our data show that IEG is far more likely to downgrade than upgrade World Bank project evaluations—nearly 30% of projects are downgraded, while fewer than 3% receive an upgrade.

To examine potential variation in how accommodation dynamics manifest across different evaluation modalities, we disaggregate IEG reviews by evaluation type. As noted earlier, all projects in our sample receive an ICRR—an expedited, desk-based validation of the self-assessed ICR. In contrast, a non-random subset of projects is selected for a PPAR, which constitutes a more in-depth and independent evaluation based on additional evidence, field visits, and broader stakeholder interviews. These PPARs are not only significantly more detailed—often exceeding 50 pages—but also more likely to challenge the original ratings: 49% of PPARs result in downgrades, compared to only 26% of ICRRs. The field-based nature of PPARs brings evaluators into closer contact with borrower governments and project sites, introducing a higher degree of scrutiny and potential political sensitivity.

This distinction in evaluation intensity is not just procedural—it shapes the likelihood that political considerations become visible in review outcomes. ICRRs rely heavily on the narrative and data provided by Bank staff and often involve limited engagement beyond the final task team leader. By contrast, PPARs introduce new, independent evidence and triangulation, raising the bar for accountability and creating more space for divergence from original self-assessments. As a result, separating the two streams is both conceptually sound and empirically necessary when analyzing whether accommodation dynamics persist under more rigorous oversight.

Table 4 presents the empirical results from models estimated separately by evaluation type. In the ICRR subsample (Columns 1–3), we find no significant association between AIIB founding membership and the likelihood of receiving a downgrade. This null result, however, should not be interpreted as the absence of accommodation. As shown earlier,

29. As a robustness check, we also estimate models using the IEG evaluation score directly as the dependent variable while controlling for the original ICR rating; results reported in Table A.1 are consistent with the patterns discussed here.

ICR ratings for AIIB founders tend to be inflated (Table 1); if the IEG were fully independent and unaffected by political considerations, one might expect more frequent downgrades for these projects. The absence of such a pattern suggests that accommodation may influence not just the initial ICR stage but also the subsequent IEG validation.

Analysis of the PPAR subsample (Columns 4-6) yields even stronger evidence of political bias in favor of the AIIB founders. In the top panel of Table 4, which uses a binary indicator for downgrades across the satisfactory/unsatisfactory threshold, the estimated coefficients for AIIB founders are consistently negative, though only statistically significant in Column 5. In contrast, the bottom panel—which uses a more inclusive dummy variable capturing any reduction in evaluation score—shows negative and statistically significant coefficients across all specifications. Given that PPARs exhibit a higher overall baseline downgrade rate, the muted downgrade likelihood for AIIB founders is particularly striking. This pattern reinforces our broader interpretation: even under conditions of heightened evaluative rigor, the World Bank appears to systematically accommodate borrowers central to the AIIB’s formation. The upshot is that when subject to the more rigorous and independent scrutiny of a PPAR, projects associated with AIIB founding members are less likely to be downgraded.

These findings carry important implications. As highlighted in Interview A, a downgrade—especially one that crosses the satisfactory/unsatisfactory threshold—can shape internal lending decisions and the stance of the Executive Board. The lower downgrade probability observed for AIIB founders, particularly in PPARs, suggests that strategic accommodation persists even at the final and most independent stage of the project evaluation process.

Table 4: AIIB Founders and IEG Downgrades

	ICRR Validation			PPAR Evaluation		
	IEG Downgrade (Binary)					
	(1)	(2)	(3)	(4)	(5)	(6)
AIIB Founder x Post-2016	-0.008 (0.030)	-0.007 (0.029)	0.001 (0.026)	-0.196 (0.116)	-0.226* (0.121)	-0.219 (0.133)
R2	0.092	0.097	0.092	0.415	0.453	0.461
R2 Adj.	0.038	0.039	0.033	0.198	0.212	0.202
IEG Downgrade (Dummy)						
	(1)	(2)	(3)	(4)	(5)	(6)
AIIB Founder x Post-2016	-0.022 (0.033)	-0.016 (0.033)	0.011 (0.037)	-0.372** (0.134)	-0.385*** (0.115)	-0.373*** (0.119)
R2	0.107	0.113	0.112	0.426	0.453	0.484
R2 Adj.	0.053	0.055	0.054	0.213	0.212	0.236
Observations	2673	2673	2540	348	348	343
Countries	138	138	122	82	82	79
AIIB Founders	24	24	24	22	22	22
Country Fixed Effects	✓	✓	✓	✓	✓	✓
Evaluation Year Fixed Effects	✓	✓	✓	✓	✓	✓
Sector Fixed Effects		✓	✓		✓	✓
Project-level Controls		✓	✓		✓	✓
Country-Year Controls			✓			✓

Notes: Results from ordinary least squares regression. Robust standard errors clustered at country and evaluation-year levels reported in parentheses.

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

7.6 Additional Evidence: PPAR Selection and Country-Level Evaluations

To further probe the durability and scope of the accommodation logic, we conduct two supplementary analyses: one at the level of PPAR selection and another at the level of country engagement. Both tests help clarify whether the observed evaluative bias is confined to project-level scoring or embedded more deeply in the institutional review process.

We first examine whether projects from AIIB founding members are less likely to be selected for PPAR evaluation. This is a crucial consideration because PPARs, unlike ICRRs, represent

the most rigorous and independent form of project review. However, the selection process is non-random, and evaluations typically occur years after project closure. Because our sample ends in 2020, some eligible projects may not yet have undergone PPAR review, introducing potential right-censoring. We therefore interpret these results with appropriate caution.

As shown in Table A.4, the estimated coefficients on the AIIB Founder \times Post-2016 interaction are negative across all specifications, though not statistically significant. These findings suggest that AIIB founders may be modestly less likely to face the more stringent scrutiny of a PPAR. While the evidence is not conclusive, the consistent direction of the coefficients aligns with our broader findings: AIIB founders' projects are both less likely to be downgraded under PPAR and, potentially, less likely to be subjected to PPAR review in the first place.

We also investigate whether the accommodation logic extends beyond individual projects to shape country-level strategic assessments. Specifically, we examine IEG's Completion and Learning Review Validations (CLRVs), which evaluate the outcomes of Country Partnership Frameworks (CPFs). As shown in Table A.5, we find that AIIB founding members are more likely to receive favorable CLRV scores.

This result is consistent with our theoretical expectations. Project evaluations inform CPF cycles, which in turn guide country strategies, lending allocations, and Board-level discussions. The fact that AIIB founders receive more positive CLRV ratings suggests that accommodation may operate not just at the level of individual project evaluations but also across successive layers of institutional review. This layered pattern—reduced downgrades under PPAR, lower likelihood of PPAR selection, and more favorable country strategy evaluations—bolsters our argument that the World Bank adopts a systematic, strategic posture toward borrowers that are geopolitically salient in the context of China's rise and the founding of the AIIB.

8 Discussion

International relations sometimes involve subtle accommodation—payoffs that are deftly wielded to win back friends and ward off non-cooperation in ways that do not provoke

widespread contagion. The emergence of the AIIB has introduced a pivotal challenge to the World Bank, forcing it to navigate the competitive pressures of contested multilateralism. We argue that the World Bank's project evaluations represent a strategic response to the accommodation dilemma posed by the AIIB's establishment. By offering favorable evaluations to AIIB founding members, the World Bank aims to preserve cooperation gains, such as retaining key borrowers, while mitigating the reputational and contagion risks inherent in accommodating non-cooperation.

The findings of this study contribute to the broader theoretical framework of the accommodation dilemma by highlighting how evaluations—often viewed as technical instruments—can serve as a nuanced mechanism for managing institutional competition. In doing so, the paper adds to the growing body of literature on the responses of IOs to non-cooperation, offering evidence of how established institutions adapt their practices to maintain influence in an evolving global order.

Future research should expand on these preliminary findings by exploring the long-term consequences of strategic evaluation practices. Do favorable evaluations successfully retain key borrowers, or do they risk undermining the credibility of evaluation systems over time? Additionally, comparative analyses across other multilateral development banks could provide valuable insights into how IOs balance flexibility and institutional integrity in the face of competition.

A vibrant literature in political science examines how the policy instruments at the disposal of IOs can be used to pursue a range of goals—from the technocratic to the political. Most of this research understandably focuses on the kinds of actions that grab international attention. When it comes to the World Bank, scholars have long studied its loans and conditionality as well as the power governments hold at the institution's executive board. More recent research has begun to examine more subtle policy tools, such as project evaluation ratings. In navigating the fine line between accommodation and credibility, the World Bank offers a compelling case for understanding how IOs leverage seemingly neutral tools, like project evaluations, to address the challenges of contested multilateralism and preserve their standing in the international system.

References

Ahluwalia, Montek Singh, Lawrence Summers, Andrés Velasco, Nancy Birdsall, and Scott Morris. 2016. “Multilateral development banking for this century’s development challenges: five recommendations to shareholders of the old and new Multilateral Development Banks.”

Apolinario Junior, Laerte, and Felipe Jukemura. 2022. “A comparative analysis of the environmental and social policies of the AIIB and World Bank.” *Global Policy* 13 (5): 694–709.

Bailey, Michael A, Anton Strezhnev, and Erik Voeten. 2017. “Estimating dynamic state preferences from United Nations voting data.” *Journal of Conflict Resolution* 61 (2): 430–456.

Bulman, David, Walter Kolkma, and Aart Kraay. 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, no. 3 (September 1, 2017): 335–363.

Cameron, A Colin, Jonah B Gelbach, and Douglas L Miller. 2011. “Robust inference with multiway clustering.” *Journal of Business & Economic Statistics* 29 (2): 238–249.

Clark, Richard, and Lindsay Dolan. 2021. “Pleasing the principal: US influence in World Bank policymaking.” *American Journal of Political Science* 65 (1): 36–51.

Cooley, Alexander, and Daniel Nexon. 2020. *Exit from hegemony: The unraveling of the American global order*. New York: Oxford University Press.

Coppedge, Michael, John Gerring, Carl Henrik Knutsen, Staffan I. Lindberg, Jan Teorell, David Altman, Fabio Angiolillo, et al. 2024. *V-Dem Country-Year Dataset v14*.

Cruz, Cesi, and Christina J Schneider. 2017. “Foreign aid and undeserved credit claiming.” *American Journal of Political Science* 61 (2): 396–408.

Daßler, Benjamin, Tim Heinkelmann-Wild, and Andreas Kruck. 2024. “How negative institutional power moderates contestation: Explaining dissatisfied powers’ strategies towards international institutions.” *The Review of International Organizations*, 1–38.

Dellmuth, Lisa, and Stefanie Walter. 2024. *How Domestic and International Actors Respond to Non-cooperation*. Working Paper.

———. 2025. *Responding to Non-Cooperation in Global Governance*. Working Paper.

Denizer, Cevdet, Daniel Kaufmann, and Aart Kraay. 2013. “Good countries or good projects? Macro and micro correlates of World Bank project performance.” *Journal of Development Economics* 105 (November): 288–302.

Desai, Raj M, and James Raymond Vreeland. 2015. "How to stop worrying and love the Asian Infrastructure Investment Bank." *Washington Post* April 6.

Dreher, Axel, Stephan Klasen, James Raymond Vreeland, and Eric Werker. 2013. "The Costs of Favoritism: Is Politically Driven Aid Less Effective?" Publisher: The University of Chicago Press, *Economic Development and Cultural Change* 62, no. 1 (October): 157–191.

Dreher, Axel, Jan-Egbert Sturm, and James Vreeland. 2009. "Development Aid and International Politics: Does Membership on the UN Security Council Influence World Bank Decisions?" *Journal of Development Economics* 88 (1): 1–18.

Fleck, Robert K, and Christopher Kilby. 2006. "World Bank Independence: A Model and Statistical Analysis of US Influence." *Review of Development Economics* 10 (2): 224–240.

Freeman, Carla P. 2019. "Constructive engagement? The US and the AIIB." *Global Policy* 10 (4): 667–676.

Girod, Desha M., and Jennifer L. Tobin. 2016. "Take the Money and Run: The Determinants of Compliance with Aid Agreements." *International Organization* 70, no. 1 (January): 209–239.

Gray, Julia. 2018. "Life, death, or zombie? The vitality of international organizations." *International Studies Quarterly* 62 (1): 1–13.

———. 2024. "The life cycle of international cooperation: Introduction to the special issue." *The Review of International Organizations*, 1–24.

Heinzel, Mirko. 2022. "International bureaucrats and organizational performance. Country-specific knowledge and sectoral knowledge in World Bank projects." *International Studies Quarterly* 66 (2): sqac013.

Heinzel, Mirko, Ben Cormier, and Bernhard Reinsberg. 2023. "Earmarked Funding and the Control–Performance Trade-Off in International Development Organizations." *International Organization* 77, no. 2 (February): 475–495.

Heinzel, Mirko, and Andrea Liese. 2021. "Managing performance and winning trust: how World Bank staff shape recipient performance." *The Review of International Organizations* 16, no. 3 (July 1, 2021): 625–653.

Heldt, Eugénia C, and Henning Schmidtke. 2019. "Explaining coherence in international regime complexes: How the World Bank shapes the field of multilateral development finance." *Review of International Political Economy* 26 (6): 1160–1186.

Hernandez, Diego. 2017. "Are "new" donors challenging World Bank conditionality?" *World Development* 96:529–549.

Hobolt, Sara B, Sebastian Adrian Popa, Wouter Van der Brug, and Hermann Schmitt. 2022.

“The Brexit deterrent? How member state exit shapes public support for the European Union.” *European Union Politics* 23 (1): 100–119.

Honig, Dan. 2019. “When Reporting Undermines Performance: The Costs of Politically Constrained Organizational Autonomy in Foreign Aid Implementation.” *International Organization* 73, no. 1 (January): 171–201.

Honig, Dan, Ranjit Lall, and Bradley C. Parks. 2023. “When Does Transparency Improve Institutional Performance? Evidence from 20,000 Projects in 183 Countries.” *Eprint*: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/ajps.12698>, *American Journal of Political Science* 67 (4): 1096–1116.

Humphrey, Chris. 2015. *Challenges and Opportunities for Multilateral Development Banks in 21st Century Infrastructure Finance*. Infrastructure Finance in the Developing World Working Paper Series.

———. 2022. *Financing the Future: Multilateral Development Banks in the Changing World Order of the 21st Century*. Oxford University Press, December.

IEG (World Bank Independent Evaluation Group). 2014. *Guidelines for Reviewing World Bank Implementation Completion and Results Reports: A Manual for Evaluators*. https://ieg.worldbankgroup.org/sites/default/files/Data/ICRR_EvaluatorManualAugust2014.pdf.

———. 2015. *World Bank Project Performance Ratings - Codebook*. https://ieg.worldbankgroup.org/sites/default/files/Data/reports/ieg-wb-project-performance-ratings-codebook_092015.pdf.

———. 2019. *2019 IEG Client Survey Final Report*. https://ieg.worldbankgroup.org/sites/default/files/Data/ClientSurveyReports/IEG_Client_Survey_Report_2019.pdf.

Jankauskas, Vytautas, and Steffen Eckhard. 2023. *The Politics of Evaluation in International Organizations*. Oxford University Press, May.

Jurado, Ignacio, Sandra León, and Stefanie Walter. 2022a. “Brexit dilemmas: Shaping post-withdrawal relations with a leaving state.” *International Organization* 76 (2): 273–304.

———. 2022b. “Brexit Dilemmas: Shaping Postwithdrawal Relations with a Leaving State.” *International Organization* 76 (2): 273–304.

Katada, Saori N. 2016. “At the crossroads: the TPP, AIIB, and Japan’s foreign economic strategy.”

Kellerman, Miles. 2019. “The Proliferation of Multilateral Development Banks.” *Review of International Organizations* 14 (1): 107–145.

Kersting, Erasmus, and Christopher Kilby. 2016. “With a Little Help From My Friends: Global Electioneering and World Bank Lending.” *Journal of Development Economics* 121:153–165.

Kilby, Christopher. 2015. "Assessing the impact of World Bank preparation on project outcomes." *Journal of Development Economics* 115 (July): 111–123.

Kilby, Christopher, and Katharina Michaelowa. 2019. "What influences World Bank project evaluations?" *Lessons on foreign aid and economic development: Micro and macro perspectives*, 109–150.

Kim, Soo Yeon, and Jesslene Lee. 2020. *Gaining Ground, Gaining Influence? Vote Shares and Power in the AIIB*. Working Paper.

Lipscy, Phillip Y. 2017. *Renegotiating the world order: Institutional change in international relations*. Cambridge University Press.

Malet, Giorgio, and Stefanie Walter. 2024. "Have your cake and eat it, too? Switzerland and the feasibility of differentiated integration after Brexit." *West European Politics* 47 (5): 1150–1179.

Malik, Rabia, and Randall W Stone. 2018. "Corporate influence in World Bank lending." *The Journal of Politics* 80 (1): 103–118.

Michaelowa, Axel, and Katharina Michaelowa. 2011. "Coding error or statistical embellishment? The political economy of reporting climate aid." *World Development* 39 (11): 2010–2020.

Michaelowa, Katharina, and Axel Borrman. 2006. "Evaluation bias and incentive structures in bi-and multilateral aid agencies." *Review of development economics* 10 (2): 313–329.

Morse, Julia C, and Robert O Keohane. 2014. "Contested multilateralism." *The Review of international organizations* 9:385–412.

Paradise, James F. 2019. "China's quest for global economic governance reform." *Journal of Chinese Political Science* 24 (3): 471–493.

Park, Susan. 2007. "The World Bank Group: Championing Sustainable Development Norms?" *Global Governance* 13 (4): 535–556.

Pemstein, Daniel, Kyle L. Marquardt, Eitan Tzelgov, Yi-ting Wang, Juraj Medzihorsky, Joshua Krusell, Farhad Miri, and Johannes von Römer. 2024. "The V-Dem Measurement Model: Latent Variable Analysis for Cross-National and Cross-Temporal Expert-Coded Data." *Varieties of Democracy Institute Working Paper* 21 (9th Ed).

Pleger, Lyn, Fritz Sager, Michael Morris, Wolfgang Meyer, and Reinhard Stockmann. 2017. "Are some countries more prone to pressure evaluators than others? Comparing findings from the United States, United Kingdom, Germany, and Switzerland." *American Journal of Evaluation* 38 (3): 315–328.

Pratt, Tyler. 2021. "Angling for influence: Institutional proliferation in development banking." *International Studies Quarterly* 65 (1): 95–108.

Qian, Jing, James Vreeland, and Jianzhi Zhao. 2023. “The Impact of China’s AIIB on the World Bank.” *International Organization* 77 (1): 217–237.

———. 2025. “In-Group Punishment in International Relations: U.S. Reactions to the Founding of China’s AIIB.” *Review of International Political Economy* 32 (4): 1073–1099.

Ravallion, Martin. 2016. “The World Bank: Why it is still needed and why it still disappoints.” *Journal of Economic Perspectives* 30 (1): 77–94.

Rodrigues Vieira, Vinicius. 2018. “Who Joins Counter-Hegemonic Igos? Early and Late Members of the China-Led Asian Infrastructure Investment Bank.” *Research & Politics* 5 (2): 2053168018770031.

Schmidt, Averell. 2024. “Treaty withdrawal and the development of international law.” *The Review of International Organizations*, 1–24.

———. 2025. “Damaged relations: How treaty withdrawal impacts international cooperation.” *American Journal of Political Science* 69 (1): 223–239.

Shin, Wonkyu, Youngwan Kim, and Hyuk-Sang Sohn. 2017. “Do Different Implementing Partnerships Lead to Different Project Outcomes? Evidence from the World Bank Project-Level Evaluation Data.” *World Development* 95 (July): 268–284.

Von Borzyskowski, Inken, and Felicity Vabulas. 2025. *Exit from International Organizations: Costly Negotiation for Institutional Change*. New York: Cambridge University Press.

Walter, Stefanie. 2020. *The mass politics of international disintegration*, CIS Working Paper 105. June. <https://doi.org/10.5167/uzh-188107>.

———. 2021a. “Brexit Domino? The Political Contagion Effects of Voter-endorsed Withdrawals from International Institutions.” *Comparative Political Studies* 54 (13): 2382–2415.

———. 2021b. “EU-27 Public opinion on Brexit.” *JCMS: Journal of Common Market Studies* 59 (3): 569–588.

———. 2023. “The Accommodation Dilemma: Balancing reputational and material concerns when responding to non-cooperative behavior.”

Walter, Stefanie, and Nicole Plotke-Scherly. 2024. *Responding to Referendum-Endorsed Challenges to Existing International Institutions*. Working Paper.

———. 2025. “Responding to Unilateral Challenges to International Institutions.” *International Studies Quarterly* 69 (2): sqaf022.

Wang, Yu. 2018. “The political economy of joining the AIIB.” *The Chinese Journal of International Politics* 11 (2): 105–130.

Watkins, Mitchell. 2022. "Undermining conditionality? The effect of Chinese development assistance on compliance with World Bank project agreements." *The Review of International Organizations* 17, no. 4 (October 1, 2022): 667–690.

Weaver, Catherine. 2008. *Hypocrisy Trap: The World Bank and the Poverty of Reform*. Princeton, NJ: Princeton University Press.

Wergin, Jon F. 1976. "The evaluation of organizational policy making: A political model." *Review of Educational Research* 46 (1): 75–115.

Wildavsky, Aaron. 1972. "The self-evaluating organization." *Public administration review*, 509–520.

Winters, Matthew S. 2019. "Too Many Cooks in the Kitchen? The Division of Financing in World Bank Projects and Project Performance." *Politics and Governance* 7, no. 2 (June 5, 2019): 117–126.

Wood, Terence, and Stephen Howes. 2025. "The impact of independence on reported aid performance." *World Development* 186:106829.

World Bank. 2011. *Implementation Completion and Results Report Guidelines*. Washington, DC. <http://documents.worldbank.org/curated/en/375311468326426795/Implementation-completion-and-results-report-guidelines>.

———. 2014. *World Bank Group Directive: Country Engagement*. <https://thedocs.worldbank.org/en/doc/d832750be4bff91a3c86b21e302852ee-0290032023/original/World-Bank-Group-Directive-Country-Engagement.pdf>.

———. 2018. *IDA17 Retrospective – Maximizing Development Impact* [in English]. Covers achievements during the IDA17 period (July 1, 2014 to June 30, 2017). Washington, D.C.: International Development Association, World Bank Group. <https://ida.worldbank.org/en/replenishments/ida17-replenishment/ida17-retrospective-maximizing-development-impact>.

———. 2021. *IDA20 - Building Back Better from the Crisis: Toward a Green, Resilient and Inclusive Future* [in English]. Washington, D.C.: World Bank Group. <http://documents.worldbank.org/curated/en/163861645554924417>.

———. 2024. *The World Bank Annual Report 2024: A Better Bank for a Better World*. Washington, DC.

Xiao, Ren. 2015. "A reform-minded status quo power? China, the G20, and reform of the international financial system." *Third World Quarterly* 36 (11): 2023–2043.

Yang, Hai, and Baldwin Van Gorp. 2019. "Framing the Asian infrastructure investment bank: a qualitative analysis of the political debate and media coverage on a China-led multilateral institution." *The Pacific Review* 32 (4): 603–634.

Zaccaria, Giuseppe. 2024. "Let's be friends not foes: an assessment of the strategic co-evolution of the World Bank and the AIIB in the face of institutional overlap." In *The Elgar Companion to the World Bank*, 166–177. Edward Elgar Publishing.

Zhao, Jianzhi, Yannan Gou, and Wanying Li. 2019. "A New Model of Multilateral Development Bank: A Comparative Study of Road Projects by the AIIB and ADB." *Journal of Chinese Political Science* 24 (2): 267–288.

Appendix

A Robustness Checks

Table A.1: AIIB Founders and IEG Validation

	ICRR Validation			PPAR Evaluation		
	IEG Evaluation (Binary)					
	(1)	(2)	(3)	(4)	(5)	(6)
AIIB Founder x Post-2016	0.014 (0.032)	0.013 (0.032)	0.006 (0.030)	0.195* (0.107)	0.215* (0.115)	0.204 (0.124)
ICR Evaluation	0.849*** (0.017)	0.847*** (0.018)	0.849*** (0.018)	0.613*** (0.076)	0.601*** (0.081)	0.612*** (0.080)
R2	0.596	0.598	0.608	0.504	0.538	0.550
R2 Adj.	0.572	0.572	0.583	0.317	0.332	0.331
	IEG Evaluation (Scale)					
	(1)	(2)	(3)	(4)	(5)	(6)
AIIB Founder x Post-2016	0.042 (0.043)	0.034 (0.042)	-0.001 (0.048)	0.449*** (0.123)	0.441*** (0.129)	0.452** (0.170)
ICR Evaluation	0.866*** (0.014)	0.864*** (0.014)	0.864*** (0.014)	0.684*** (0.028)	0.669*** (0.034)	0.647*** (0.033)
R2	0.723	0.725	0.729	0.664	0.671	0.683
R2 Adj.	0.706	0.707	0.711	0.538	0.524	0.529
Observations	2673	2673	2540	348	348	343
Countries	138	138	122	82	82	79
AIIB Founders	24	24	24	22	22	22
Country Fixed Effects	✓	✓	✓	✓	✓	✓
Evaluation Year Fixed Effects	✓	✓	✓	✓	✓	✓
Sector Fixed Effects		✓	✓		✓	✓
Project-level Controls		✓	✓		✓	✓
Country-Year Controls			✓			✓

Notes: Results from ordinary least squares regression. Robust standard errors clustered at country and evaluation-year levels reported in parentheses.

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

Table A.2: Formal Membership: AIIB Founder and ICR Evaluation

	ICR Evaluation (Binary)			ICR Evaluation (Scale)		
	(1)	(2)	(3)	(4)	(5)	(6)
AIIB Founder (Formal Membership)	0.066*	0.065**	0.067*	0.178*	0.174**	0.182**
	(0.036)	(0.030)	(0.034)	(0.089)	(0.075)	(0.066)
Project cost (log)		0.032***	0.033***		0.107***	0.112***
		(0.007)	(0.007)		(0.020)	(0.019)
GDP per capita (log)			0.009			0.059
			(0.064)			(0.094)
Population (log)			0.146			0.405
			(0.148)			(0.406)
FDI inflow (% GDP)			-0.001			-0.002
			(0.001)			(0.002)
Natural resource rent (% GDP)			0.000			0.004
			(0.003)			(0.008)
Corruption			-0.195			-0.352
			(0.115)			(0.254)
Democracy			-0.173			-0.557*
			(0.131)			(0.263)
UNGA voting (Ideal Point distance from US)			-0.010			-0.026
			(0.043)			(0.089)
Temporary UNSC member			0.044			0.136**
			(0.026)			(0.047)
Observations	3192	3192	3049	3192	3192	3049
Countries	138	138	122	138	138	122
AIIB Founders	26	26	26	26	26	26
R2	0.097	0.111	0.107	0.143	0.166	0.167
R2 Adj.	0.053	0.063	0.059	0.101	0.121	0.122
Country Fixed Effects	✓	✓	✓	✓	✓	✓
Evaluation Year Fixed Effects	✓	✓	✓	✓	✓	✓
Sector Fixed Effects	✓	✓		✓	✓	✓

Notes: Results from ordinary least squares regression. Robust standard errors clustered at country and evaluation-year levels reported in parentheses.

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

Table A.3: Founder vs. Non-Founder: AIIB Founder and ICR Evaluation

	ICR Evaluation (Binary)			ICR Evaluation (Scale)		
	(1)	(2)	(3)	(4)	(5)	(6)
AIIB Founder (Formal Membership)	0.064 (0.040)	0.063* (0.034)	0.060 (0.036)	0.167 (0.098)	0.161* (0.084)	0.163** (0.074)
AIIB Non-Founder (Formal Membership)	-0.021 (0.086)	-0.031 (0.088)	-0.079 (0.102)	-0.120 (0.172)	-0.139 (0.178)	-0.237 (0.203)
Project cost (log)		0.032*** (0.007)	0.033*** (0.007)		0.108*** (0.019)	0.113*** (0.019)
GDP per capita (log)			0.012 (0.063)			0.066 (0.091)
Population (log)			0.158 (0.149)			0.441 (0.409)
FDI inflow (% GDP)			-0.001 (0.001)			-0.002 (0.002)
Natural resource rent (% GDP)			0.000 (0.003)			0.003 (0.010)
Corruption			-0.199 (0.113)			-0.363 (0.247)
Democracy			-0.175 (0.130)			-0.564* (0.261)
UNGA voting (Ideal Point distance from US)			-0.010 (0.043)			-0.025 (0.089)
Temporary UNSC member			0.045 (0.027)			0.140** (0.048)
Observations	3192	3192	3049	3192	3192	3049
Countries	138	138	122	138	138	122
AIIB Founders	26	26	26	26	26	26
AIIB Non-Founders	15	15	12	15	15	12
R2	0.097	0.111	0.108	0.144	0.166	0.167
R2 Adj.	0.052	0.063	0.059	0.101	0.121	0.122
Country Fixed Effects	✓	✓	✓	✓	✓	✓
Evaluation Year Fixed Effects	✓	✓	✓	✓	✓	✓
Sector Fixed Effects	✓	✓		✓	✓	✓

Notes: Results from ordinary least squares regression. Robust standard errors clustered at country and evaluation-year levels reported in parentheses.

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

Table A.4: AIIB Founders and IEG PPAR Selection

	PPAR Selection		
	(1)	(2)	(3)
AIIB Founder x Post-2016	-0.039 (0.033)	-0.042 (0.033)	-0.034 (0.035)
Project cost (log)		0.012 (0.009)	0.013 (0.009)
ICR outcome (binary)		0.053** (0.018)	0.055** (0.019)
ICRR downgrade (dummy)		0.015 (0.010)	0.016 (0.010)
ICR quality (binary)		0.005 (0.025)	0.004 (0.027)
June ICR		0.014 (0.016)	0.016 (0.017)
GDP per capita (log)			0.017 (0.049)
Population (log)			0.047 (0.109)
FDI inflow (% GDP)			0.000 (0.002)
Natural resource rent (% GDP)			-0.001 (0.001)
Corruption			-0.098 (0.070)
Democracy			-0.001 (0.098)
UNGA voting (Ideal Point distance from US)			0.033 (0.042)
Temporary UNSC member			0.035 (0.037)
Observations	3176	3157	3014
Countries	138	138	122
AIIB Founders	26	26	26
R2	0.100	0.110	0.109
R2 Adj.	0.055	0.060	0.060
Country Fixed Effects	✓	✓	✓
Evaluation Year Fixed Effects	✓	✓	✓
Sector Fixed Effects		✓	✓

Notes: Results from ordinary least squares regression. Robust standard errors clustered at country and evaluation-year levels reported in parentheses.

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

Table A.5: AIIB Founders and IEG CLRV Evaluation

	IEG CLRV Evaluation (Binary)			
	(1)	(2)	(3)	(4)
AIIB Founder x Post-2016	0.275** (0.112)	0.166 (0.118)	0.213* (0.111)	0.183 (0.123)
RMSE	0.46	0.44	0.37	0.43
Log Likelihood	-166.8	-158.4	-184.0	-146.2
	IEG CLRV Evaluation (Scale)			
	(1)	(2)	(3)	(4)
AIIB Founder x Post-2016	0.491*** (0.187)	0.326* (0.196)	0.354* (0.187)	0.328* (0.185)
RMSE	0.71	0.68	0.54	0.66
Log Likelihood	-281.8	-272.5	-289.6	-249.8
Observations	263	263	263	250
Countries	109	109	109	99
AIIB Founders	21	21	21	21
Review Year Fixed Effects	✓	✓	✓	✓
Region Fixed Effects		✓		
Country Random Effects			✓	
Controls				✓

Notes: Results from ordinary least squares regression. Robust standard errors clustered at country level reported in parentheses. Control variables include log GDP per capita, log total population, and the V-Dem polyarchy index, all measured at the start of the review window; as well as temporary UN Security Council membership and UNGA ideal point distance to the United States, both measured in the year of review.

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