COUNTRY SELECTIVITY IN AID ALLOCATION: EVIDENCE ON NEED AND EFFECTIVENESS

by
Jake Grover

A thesis submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of International Affairs

Baltimore, Maryland
May 2023

Abstract

Over $200 billion was spent in foreign assistance in 2022, yet most donors do not have explicit criteria for allocating their resources. The misallocation of foreign aid resources can create huge inefficiencies that potentially stifle its effectiveness. This study produces evidence for the optimal allocation of foreign assistance across the dimensions of country need and potential effectiveness for donors that seek to maximize their impact on poverty reduction through economic growth.

My research uses quantitative methods to examine the two most relevant allocation factors for the goal of poverty reduction through growth: need and effectiveness. The first chapter reviews the aid allocation literature and proposes a conceptual framework to guide the rest of the analysis.

The second chapter explores a needs-based approach across the dimensions of the development challenge and the resources available. The two different components of country need both suggest a strong focus on allocating assistance towards the poorest countries.

The third chapter examines the differential effectiveness of aid at a macro level for different criteria employed by performance-based approaches – i.e., is foreign aid more effective in promoting economic growth in better-governed and more democratic countries? I find that the aid-growth relationship is much stronger for worse-governed and less democratic countries.

The fourth chapter exploits micro data to examine differences in project-level outcomes for both needs-based and performance-based aid allocation criteria. In find that good governance is the most important country-level factor, followed by higher average income, corruption is insignificant, and democracy may be a detriment to achieving project outcomes.

ABSTRACT

The fifth chapter concludes by comparing the current allocation of assistance to more evidence-based optimal allocation models, and I find that too much assistance goes to richer countries and strategic partners. I conclude by providing recommendations for improving the effectiveness of foreign assistance through revised allocation criteria based on the findings of the preceding chapters.

Primary reader and advisor: Sam Asher

Secondary reader: David Bulman

External reader: Stephen O'Connell

Preface

"Since there was only so much aid to go around, the people who paid the price for the large flows of aid to environments in which it was ineffective were those living in the better policy environments who could otherwise have been lifted out of poverty."

- Paul Collier, April 1999

Contents

List of Tables

List of Figures

Chapter 1. Background and Conceptual Framework

Over $200 billion was spent in foreign assistance in 2022 by Development Assistance Committee (DAC) countries,[1] yet most donors do not have explicit criteria for allocating their resources.[2] At the same time, the international community committed to achieving the Sustainable Development Goals (SDGs) by 2030, including ending poverty in all its forms, and this implies resource needs that far outstrip the available resources. Beyond accountability to taxpayers for providing these resources,[3] the massive development financing gap magnifies the need to stretch the development impact of every aid dollar. The misallocation of foreign aid resources can create huge inefficiencies that potentially stifle its effectiveness. This study produces evidence for the optimal allocation of foreign assistance across the dimensions of country need and potential effectiveness for donors that seek to maximize their impact on poverty reduction through economic growth.

The rest of the study proceeds as follows. This chapter reviews the aid allocation literature and proposes a conceptual framework to guide the rest of the analysis. The second chapter explores a needs-based approach across the dimensions of the development challenge and the resources available. The third chapter examines the differential effectiveness of aid at a macro level for different criteria employed by performance-based approaches. The fourth chapter exploits micro data to examine differences in project-level outcomes for both needs-based and performance-based aid allocation criteria. The fifth chapter concludes by comparing the current allocation of assistance to more evidence-based optimal

---

[1] The Organisation for Economic Cooperation and Development's (OECD) Development Assistance Committee is an international forum consisting of 31 of the leading bilateral donors.

[2] This is despite numerous advantages: greater transparency and accountability and greater consistency, predictability, and coordination of funding decisions (Ottersen *et al* 2017). Anderson (2008) recommends that donors state more explicitly the principles guiding their aid allocations.

[3] Anderson (2008) makes three arguments in favor of performance-based allocations: efficiency in terms of poverty reduction effectiveness; providing an incentive to countries through a signaling mechanism; and transparency being publicly available and easy to understand.

allocation models and provides recommendations for improving the effectiveness of foreign assistance through revised allocation criteria based on the findings of the preceding chapters.

Historical background

The international community's focus on poverty reduction has waxed and waned over the last 50 years (Easterly 2007). In 1977, World Bank President McNamara declared that "the rich and the powerful have a moral obligation to assist the poor", but the focus receded in the 1980s as structural adjustment took center stage. Meanwhile, USAID emphasized a "basic human needs" approach in the late 1970s and early 1980s (Crosswell 2015). The World Bank then reviewed its poverty focus in the 1990s, and this culminated in the Millennium Development Goals (MDGs), which were launched in 2000 with eight poverty reduction targets through 2015. The UK enshrined the goal of poverty reduction into law in 2002, and the Millennium Challenge Corporation (MCC) was created with the mission of poverty reduction through growth in 2004. The World Bank adopted the twin goals of ending extreme poverty and promoting shared prosperity in 2013, USAID adopted the goal of ending extreme poverty (briefly) into its mission in 2014, and the MDGs evolved into the SDGs in 2015. It is now widely accepted that the over-arching goal of economic aid and development assistance is some variant of poverty reduction and/or economic growth.

The economic literature on the allocation of aid in pursuing these objectives has followed a similar trajectory. The economic literature on aid allocation can be traced back to at least Dudley and Montmarquette (1976). Cogneau and Naudet (2007) review the intellectual history and categorize the research into three phases. First, earlier studies focused on donor motivations, such as McKinlay and Little (1977), which tried to distinguish between a needs-based approach and a more geopolitically driven allocation of resources. Second, research shifted towards a greater focus on effectiveness by incorporating the quality of government institutions and economic policies (e.g., Burnside and Dollar 2000). Finally, the international community increased it focus on poverty reduction as the foremost goal of foreign aid, and this was the explicit objective of an allocation formula developed by Collier and Dollar (2002) and later attempts to establish an optimal set of criteria.

The literature can perhaps be usefully categorized into two approaches: first, a set of descriptive (or positive) studies that are backwards-looking and attempt to explain the observed allocation of assistance, typically by exploring various potential motivating allocation factors and/or the differences between donors; and second, prescriptive (or normative) studies that are more forward-looking and provide recommendations of how donors *should* allocate aid – often assessing the quality of the current aid allocation against the researcher's preferred allocation criteria. I explore these approaches in the next two sections and then suggested alternatives in a final section.

Descriptive studies

Early aid allocation studies used a framework that distinguished between need and donor self-interest. McKinlay and Little (1977) found that the United States' allocation was determined by the power structure of the Cold War-era international system with geo-political interests dominating the decision-making process with issues of effectiveness being of little concern. Similarly, Alesina and Dollar (2000) found that bilateral aid was determined by strategic interests far more than need. Historical ties and alliances were the key drivers of aid allocations, and democratization led to a dramatic increase in aid.

Burnside and Dollar (2000) changed the policy discussion by asserting that governance is a critical factor in determining the effectiveness of aid, because aid only results in growth in better-governed countries with prudent fiscal, monetary, and trade policies. The previous framework expanded to include effectiveness as measured by the policy performance of a particular country. Burnside and Dollar (2000) was very influential, and several donors became more selective on the basis of policy performance (Quibria 2014). For example, the World Bank instituted "*ex post* conditionality" that rewarded countries for policy performance as opposed to relying on threats to later withdraw funding if policy reform commitments were not upheld. However, the timing of this shift in allocation criteria was disputed.

Dollar and Levin (2006) found that selectivity in aid allocation was a relatively recent phenomenon. While most donors consistently supported democracies between the late 1980s and the early 2000s, donors became increasingly focused on economic governance over that period, particularly multilateral donors. Claessens *et al* (2009) confirmed that shift towards greater country selectivity. Despite significant heterogeneity in donor allocation models, they found that donors became both more poverty-focused and governance-focused in the post-Cold War era, while becoming less sensitive to the country's population and debt burden. McGillivray (2003b) refuted the finding that donors only started implementing a needs-based approach after the Cold War, and the author found that aid did not necessarily increase its focus on effectiveness following the Cold War, as other studies had found. Similarly, Easterly (2007) reviewed donor practices across a range of issues, such as selectivity and tied aid, and observed that needs-based selectivity is an old idea going back to at least the Pearson Commission in 1969. Most of the shift towards the poorest countries happened around 1973 – possibly in response to the McNamara vision quoted above. Further, the author found no evidence of a further shift towards greater needs-based allocations following the end of the Cold War, except for the US,[4] or for greater governance-based selectivity.

[4] Grover (2009) found that US motivations shifted towards commercial interests in the 1990s but then back towards geopolitical interests post-9/11 as well as a greater needs focus driven by the aid effectiveness agenda.

As the policy discussion of allocation criteria progressed, Anderson (2008) explored other potential needs-based allocation criteria might include non-monetary outcomes, such as the Human Development Index (HDI), as well as a country's ability to obtain revenues from alternative sources. The author found that income matters more than non-income criteria; policy performance matters more than development outcomes; population has a negative effect on aid per capita; and countries in "special circumstances" do not receive more aid when controlling for these other factors. Ottersen *et al* (2017) summarized previous by observing that GNI per capita is widely used but confirmed that bilateral donors are often influenced by a host of historical ties and political and commercial interests.

Prescriptive studies

One of the first papers to take a more prescriptive approach was Collier and Dollar (2002) who proposed a "poverty efficient" allocation of aid and compared it to the allocation of assistance at the time. They asserted that their allocation formula would have the greatest impact on poverty for any chosen measure of poverty (poverty rate, poverty gap, or severity of poverty) and the quality of governance. In comparing actual aid allocations to their poverty-efficient model, they found that aid was grossly mis-allocated, and a better allocation could roughly double the poverty impact of assistance. Collier and Dollar (2002) make a number of questionable assumptions,[5] and the paper motivated a number of similar studies that were also normative in nature.

Beynon (2003) criticized the assumptions of Collier and Dollar (2002) and pointed to a range of other factors that influence the effectiveness of aid besides policy performance, such as economic vulnerability, external shocks, conflict, and geographic factors. The author suggested donors should also

[5] Their questionable assumptions include that aid is distribution neutral, the growth elasticity of poverty is equal across countries, and that the growth elasticity of poverty is two.

be incorporating these other factors, including accounting for uneven regional trends in poverty reduction and the MDGs. Similarly, Mosley *et al* (2004) examined the direct relationship between aid and poverty reduction – thus setting aside the intermediating relationship with economic growth. The authors devised a "pro-poor public expenditure index" and find that control of corruption and pro-poor public spending are the keys to aid effectiveness. They also contest Collier and Dollar's (2002) assumption that donors cannot influence policy – rather, Mosley *et al* (2004) found that donors *can* influence pro-poor spending policy, and a "new conditionality" could provide leverage to incentivize governments to reform their spending to be more pro-poor.

Another thread of the literature explored the normative aspect of Collier and Dollar (2000) in contrasting an optimal allocation to the current allocation. For instance, McGillivray (2003b) found that all donors – even those with strong reputations for reasonable allocations of assistance – would need to make significant changes to the allocation of their assistance to approach the normative models. Bigsten *et al* (2011) reviewed European aid in light of the aid effectiveness agenda and attempted to quantify potential gains through donor reforms. They find that approximately €19 billion of the EU's €27 billion in aid would need to be re-allocated to a small set of 20 countries, and this would produce a net gain of about €7.8 billion after adjusting for the quality of governance.

Ceriani and Verme (2014) proposed a measure of the redistributive capacity that they refer to as the "income lever". This builds on the insight of Kanbur and Mukherjee (2007) that two countries with similar poverty rates should not be treated equally when one has more taxable resources for redistribution due to a more unequal distributions of income – an axiom they call a "poverty reduction failure". Building on Quiggin and Madhavan (2010) who compare the poverty gap to total resources in an economy above the poverty line, Ceriani and Verme (2014) constructed three different measures of redistributive capacity based on the necessary marginal tax rate of the non-poor, a lump-sum tax on all citizens above the poverty line plus the lump sum (so the lump sum does not push someone into poverty); and the marginal tax required of resources about some line above which citizens are considered rich. (These concepts are explored further in the next chapter.) The authors then compare their "income lever" needs-based measures against allocations based on average income and the poverty rate. Ceriani and Verme (2014) found that the actual distribution of aid does not resemble any of these criteria in practice, and the current allocation criteria are neither clear nor transparent.

Similarly, Ottersen *et al* (2017) found that quantitative data are not widely used for allocation purposes related to global health assistance, and the criteria used vary widely. Furthermore, donor understandings of the drivers of effectiveness vary – from expected impacts to policy performance to absorptive capacity –and these criteria rarely refer to the development outcome intended by the funding. The authors examined the allocation of global health assistance, and they simulated 11 different allocation criteria with very different outcomes. They find that low-income countries (LICs) benefit most from a needs-based allocation linked to domestic capacity, while upper middle-income countries (UMICs) benefit the most from criteria based on measures of inequality. Across almost all criteria, LICs get less assistance compared to a simple criterion of just average income.

Another set of related papers focuses on the trade-offs between need and effectiveness. Easterly and Pfutze (2008) drew out the tension between needs-based and governance-based criteria by noting that aid is less effective when it goes to countries ruled by corrupt dictators or relatively richer countries, yet poorer countries are more likely to be ruled by corrupt dictators. Bourguignon and Platteau (2015) asked whether a needs-based allocation of resources can negatively impact effectiveness based on a similar observation that the neediest countries are often the worst governed – thus setting up a tension in the traditional allocation model. The authors theorized on the trade-offs with the central insight being that a needs-based approach starts by serving the needs of the poorest, often in the worst governed countries, whereas a governance-based approach starts in better governed countries where there is less severe poverty. Bourguignon and Platteau (2022) later proposed an optimal allocation model recognizing the trade-off between need and effectiveness that they called the "need-adjusted aid effectiveness" approach. They recommend that donors should not consider governance in the very poorest countries but should instead look to improve policy or influence project management.

A final set of papers focused mainly on the concern of absorptive capacity, i.e., the ability of countries to use aid effectively, which often entails identifying an upper limit in terms of the amount of aid that a country can reasonably use effectively expressed as a proportion of its GDP. Early on, McGillivray and White (1995) provided three principles for a "good" allocation of aid that included absorptive capacity (in addition to need and scaling to population size). Carter (2014) questioned whether a dogmatic pursuit of a greater donor concentration of resources in the poorest countries translates into greater effectiveness. The author argued that when countries run up against the limits of their aid absorption capacity, further allocations that favor those countries may not be optimal. This implies that the weight placed on the income criteria in aid allocation formulas should be less than existing models, as this would avoid a situation whereby too much aid is directed to poor countries that are unable to absorb it effectively. To make this narrative real, Anderson *et a*l (2022) found that large aid disbursements to extremely aid-dependent countries coincide with deposits in offshore bank accounts, and the "leakage rate" was found to increase with greater aid dependency. The authors' findings suggested elite capture and corruption in aid disbursements, particularly among aid-dependent countries that are not well governed.

Alternative approaches

The logic guiding the Collier and Dollar (2000) allocation model was simple, intuitive, and compelling. Donors should allocate more resources to countries where there is greater poverty and where it can have the greatest impact. However, donors have divergent interests beyond just a needs-based approach, as established by Alesina and Dollar (2000).[6] Furthermore, the SDGs cover multi-dimensional poverty whereas Collier and Dollar (2000) are only concerned with consumption poverty. Gunning (2000) provided four objections to country selectivity: first, it denies aid to poor people living in countries with governments that perform poorly (at no fault of their own); second, countries with strong policy environments do not need aid as they can generate domestic revenues and access capital on international markets; third, governance is measured subjectively; and finally, selectivity can conflict with country ownership by forcing a reform agenda on country partners. In response to the growing literature related to needs-based and performance-based allocation systems, a number of studies proposed alternative approaches based on other criteria or considerations.

McGillivray (2006) highlighted the trade-offs between selectivity in aid allocation and the increased focus on fragile states. The author observed that fragile states receive about 43 percent of ODA yet are "under-aided" when considering the standard allocation factors of their average income, population, and CPIA scores. Furthermore, aid flows to fragile states are especially volatile, which undermines the objective of poverty reduction. McGillivray (2006) concluded that donors must resolve the coordination problem and stabilize volatile flows. That author suggested that non-traditional allocation factors, such as structural vulnerability and political stability, should also be considered – a suggestion drawn out by Guillamont (2008).

[6] However, it should be noted that the divergence between the Collier and Dollar (2000) model and the allocation at that time could not be explained by strategic interests.

Cogneau and Naudet (2006) suggested the incorporation of a non-welfarist perspective based on equalizing opportunity across countries. They incorporated issues related to fairness, including structural factors and "intangible disadvantages" that hold a country back. Bourguignon and Gunning (2020) described the Cogneau and Naudet (2007) equal opportunity approach as "compensating countries for adverse circumstances beyond their control." They did this by replacing policy performance with "circumstances" as proxied by initial characteristics. Guillamont (2008) asserted that any aid allocation formula should be based on three principles: effectiveness in contributing to its goals; an equitable allocation, however determined; and transparency in setting the criteria. With these principles in mind, Guillamont (2008) suggested an allocation model building on the "equal opportunity" model of Cogneau and Naudet (2006) to incorporate preference for countries with structural handicaps. This had an intellectual underpinning in the criteria for the UN's Least Developed Country list that includes an Economic Vulnerability Index. This would lead to greater assistance for small and fragile states.

Kappel (2017) made the case (on behalf of the Swiss government) for aid to middle-income countries (MICs) when poverty remains high. He argued that the distinction between LICs and MICs is not helpful for setting priorities and more detailed criteria than average income should be utilized. He identified a critical question: Should donors focus on poor countries or poor people? Despite his earlier assertions, he rightfully identified that richer countries have more domestic resources to address poverty and that aid allocated to MICs with higher incomes means less aid for LICs. The author suggested several criteria for allocating aid, including income, poverty and inequality, and domestic capacity with a stop-light system of high-medium-low that layers on additional, arbitrary distinctions.

Paulo *et al* (2015) examined the growing trend of thematic allocation models and sectoral earmarks, citing the example of U.S. Presidential initiatives like Feed the Future. While sectoral approaches mobilize funding and align funding to agreed international challenges like the SDGs, sectoral initiatives often bypass partner governments and increase fragmentation through parallel structures and a lack of coordination with existing activities. To the extent that these thematic allocation models persist, the authors recommended choosing a thematic area that builds on comparative advantages; employing needs-based measures tailored to that thematic area; coordinating among donors to ensure a coherent division of labor; and supporting national development strategies by emphasizing country ownership. Specific to climate changing funding, Weiler and Sanubi (2019) compared development aid to climate finance in Africa and find similar patterns across both flows, though the so-called "additionality" measures are only weak determinants of resource allocation for climate finance.

Finally, there are a set of papers that argued that some of the differential between the current allocation and optimal models may be explained by donor expectations of future poverty. For instance, Wood (2008) observed that the Collier and Dollar (2000) approach conflicted with the allocation of resources implied by the MDGs because it does not account for forward-looking expectations of poverty. The author applied a "poverty decline adjustment" to the poverty-efficient allocation formula to resolve this conflict. Barder (2009) observed the same inconsistency between allocations based on current poverty rates and expected poverty rates at some point in the future. He favored models of allocation that account for expected future poverty, as this results in allocations that are higher for country and regions that are expected to experience persistent poverty – otherwise, a significant amount of aid would be allocated to India (and previously China), which has large numbers of poor people but is growing rapidly, reducing poverty, and arguably does not need (or desire) more assistance.

Conceptual framework

Building on the extensive literature on this topic, I construct a conceptual framework that guides the analysis of the chapters to follow. As a starting point, the relevant factors and criteria for a given aid allocation model depend on a donor's goals and strategies. Arkedis (2011) stated that "Clarity about aid's purposes and the ability to make tradeoffs between the resources spent pursuing different purposes is the very definition of the 'strategic coherence'…" Likewise, Brainard (2003) asserted that it is critical to clarify the purposes for which aid is being allocated and align the principles for guiding allocation accordingly.[7] However, Berthelemy (2006) found that donors have a wide range of purposes for assistance other than poverty reduction. For instance, Grover (2009) showed that the motivations for the world's largest bilateral donor, the United States, are driven by *both* security interests and developmental need in the post-9/11 era after being mainly concerned with commercial interests in the 1990s.[8] It is important to unpack this and understand donor goals, as the considerations change dramatically when viewing the allocation criteria through a security lens or other goals aside from poverty reduction and economic growth.

While there are myriad potential goals for foreign assistance – from national security to poverty reduction to climate change and many more – for my purposes, I am focused on the goal of poverty reduction. There are numerous reasons for this: ending poverty in all its forms is SDG 1, ending extreme poverty is one of the World Bank's twin goals, "poverty reduction through economic growth" is MCC's mission, and ending extreme was previously the mission of USAID ("We partner to end extreme poverty and promote resilient democratic societies"). Those three organizations cover the largest donor (World

[7] Arkedis (2011) suggests a useful framework that distinguishes the various purposes of assistance – particularly security interests versus development goals – and catalogs other attempts to elucidate the various US foreign assistance goals.
[8] However, even where development is the primary objective, the U.S. Government foreign assistance architecture is not a monolith – rather, it is extremely fragmented – and several agencies have a narrower mandate and geographic focus, such as MCC and the Development Finance Corporation's increasing focus on LICs and LMICs

Bank), the largest bilateral donor agency (USAID), and the donor with the most data-driven and disciplined approach to country selection (MCC). There are countless other foreign aid organizations with similar goals, and it is even enshrined in law in the UK through the International Development Act.

With my goal clarified, I must then consider the relevant theory of change for achieving that goal. Some assistance providers prefer a more direct approach to poverty reduction, such as direct cash transfers (e.g., GiveDirectly) or service provision (e.g., the Global Alliance for Vaccines and Immunisation). Other actors provide direct budget support to help finance the priorities of the partner government (e.g., the International Monetary Fund). However, the World Bank (Joliffe 2014), USAID (2015), and MCC all hold up economic growth as the primary driver of poverty reduction, which is an approach strongly supported by the evidence on the drivers of poverty reduction (Dollar *et al* 2016).

The strategy for achieving poverty reduction is arguably less important than the goal itself, as the allocation factors discussed next could apply equally to a strategic approach focused on direct transfers or service provision. However, a focus on economic growth leads me to emphasize broader development progress (as proxied by average income), whereas direct transfers might be more focused on more granular measures of poverty to better target the most needy households (e.g., poverty measures disaggregated by gender or sub-nationally) and a direct service provision model might favor sectoral measures to identify which multi-dimensional poverty outcomes in which a country might be falling short (e.g., access to electricity or infant mortality). Nonetheless, while difficult to achieve – and even more difficult to claim attribution – sustained economic growth is commonly accepted as the best strategy to reduce poverty, particularly in less developed countries where average income is very low.

Since at least the *Assessing Aid* report from the World Bank (Dollar and Pritchett 1998), [9] country need and effectiveness have been considered the most relevant criteria for the allocation of foreign assistance when the primary goal is reducing poverty through economic growth (Anderson 2008 and d'Orey and Prizzon 2019). In other words, these factors *should* be the relevant allocation criteria for donors with poverty-oriented missions, such as the World Bank and MCC. If an aid provider's primary objectives are related to national security, commercial interests, or global public goods like climate change, then need and effectiveness may not be useful organizing principles or at least would be interpreted differently to give appropriate weight to the objectives that the donor is attempting to achieve (Ottersen *et al* 2017). Similarly, if a donor's motivations are more parochial than altruistic, then the criteria of need and effectiveness may matter less or not at all. This might translate into the donor directing aid towards regional spheres of influence or towards countering geopolitical rivals.

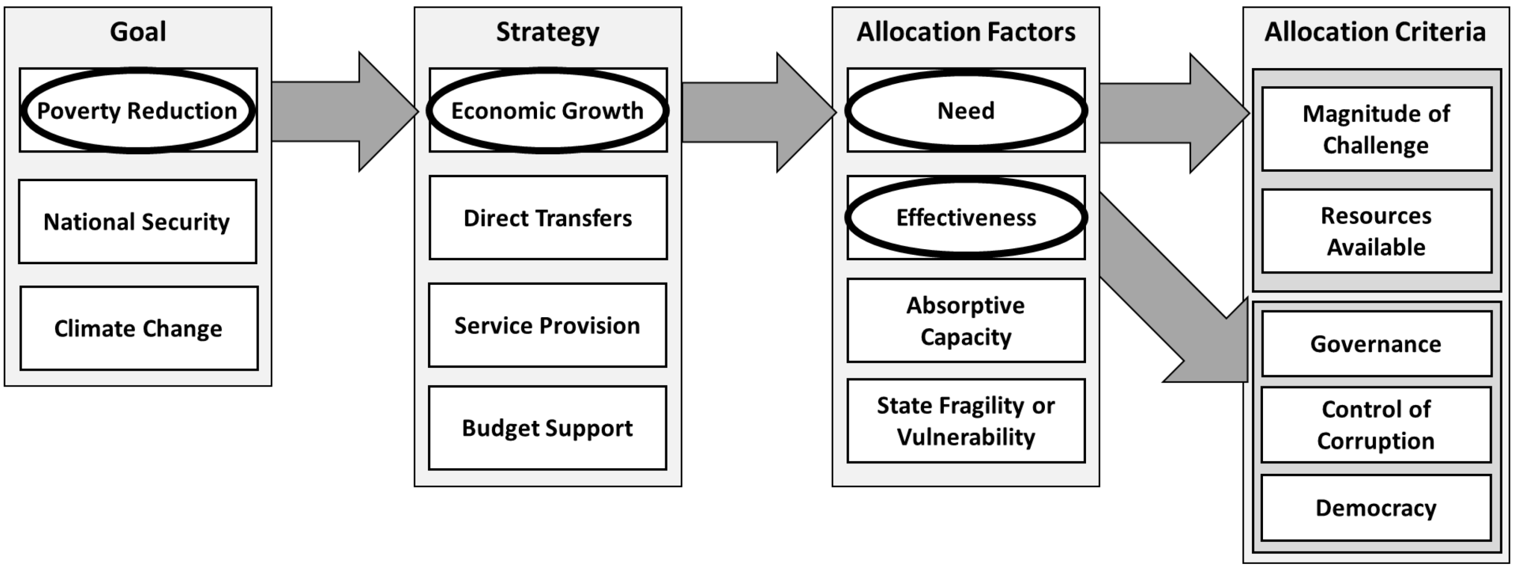
The appropriate allocation criteria then fall out of the selected allocation factors relevant to my goal and strategy. For need, I will argue in the next chapter that the most relevant criteria are the magnitude of the development challenge and the resources available. For effectiveness, the criteria include a range of governance measures on policy performance as well as democracy. I will explain the thinking behind these criteria further in the third and fourth chapters.

The conceptual framework in Figure 1 graphically depicts and summarizes how these pieces fit together. I am focused on poverty reduction as the over-arching goal. The strategic approach is to achieve poverty reduction by raising average incomes through inclusive economic growth. With this goal and strategy in mind, a donor should consider both country need and effectiveness in their allocation of assistance.

[9] The report states: "…financial assistance must be targeted more effectively to low-income countries with sound economic management… Clearly, poor countries with good policies should receive more financing than equally poor countries with weak economic management… Furthermore, much of aid continues to go to middle-income countries that do not need it."

Country need can be broken down into the scope of the development challenge and the resources available. Effectiveness can be operationalized through criteria focused on good governance and democracy. In short, a donor focused on poverty reduction through economic growth should consider two allocation factors: need and effectiveness.

Figure 1. Conceptual framework diagram of allocation factors



Research methods

My research will use quantitative methods to examine the two most relevant allocation factors for the goal of poverty reduction through growth: need and effectiveness. I conclude this chapter by previewing the methods I will use for the remaining chapters to explore this conceptual framework.

In the next chapter, I will use descriptive statistics to show that both consumption and multi-dimensional poverty are concentrated in relatively poor countries as measured by average income. This is the magnitude of the development challenge. Then I examine the redistributive capacity of developing countries by comparing their poverty gap to the taxable resources available in their economies. This is an indicative measure of the resources available for poverty reduction. When these two criteria are combined, it provides useful information for policymakers looking to allocate aid on the basis of need.

I examine two different two outcome levels when considering whether “performance-based” aid allocation leads to a more effective use of assistance. In the third chapter, I examine macro outcomes, such as economic growth. To test the relevance of the governance and democracy criteria, I employ an instrumental variables (IV) regression approach to examine whether working only with well-governed countries improves the effects of aid on growth at the macro level. In the fourth chapter, I examine micro outcomes in the form of project-level results. I utilize a fixed effects model to analyze whether the criteria related to need and effectiveness have an effect on project-level success at the micro level.

The fifth chapter concludes by incorporating the evidence presented in the previous chapters into normative models of aid allocation and then testing them. These optimal models are compared to the current allocation of assistance – both globally and for the US – and I provide recommendations to shift the current allocation closer to a more evidence-based optimal allocation.

Chapter 2. Country Need

This chapter examines the two components of country-level development need: the magnitude of the development challenge and resource availability. Given that there are huge differences in the marginal utility of income between lower-income aid recipients – LICs and LMICs – and UMICs, a significant amount of development impact could be left on the table by allocating aid to countries that need it less (Kenny 2020). Yet, donors continue to allocate huge sums of assistance to UMICs that arguably do not need it – for instance, the US spends nearly 30 percent of its assistance in UMICs, which is more than it spends in LICs. Furthermore, development policy research has defended allocations to UMICs (e.g., Alonso 2014), and MCC is currently seeking a legislative change to expand its candidate pool to include UMICs (MCC 2023). A commonly exploited stylized fact is that most poverty now exists in middle-income countries (Edward and Sumner 2013),[10] even though less than ten percent of global poverty exists in UMICs. The resultant allocation decisions come with real-world consequences. For example, the US spent more in Jordan (with an extreme poverty rate of 0.03 percent) than it did in the six countries with the highest poverty rates in the world combined, including DRC and Somalia (with 63.2 and 68.0 percent extreme poverty rates, respectively). Collier and Dollar (2002) found that if global aid were allocated according to a poverty-efficient allocation model, it could roughly double its poverty impact and benefit an additional ten million people per year.[11] In this chapter, I examine the concept of country need and explain how and why a donor should apply a needs-based approach to allocating aid. I focus specifically

[10] The analytical sleight of hand is that LMICs and UMICs are combined into one group, which makes this assertion technically true, though very misleading. Other analysts criticize the LIC/MIC distinction as not meaningful, while deliberately combining the two largest income groups (LMICs and UMICs) in terms of the breadth of average incomes included and comparing them to the smallest income group, LICs (Kappel 2017).

[11] An important caveat to note is that there are many different goals guiding the allocation of foreign assistance across donors and funding sources – from broad geopolitical and economic interests to narrow technical objectives within a sector. Furthermore, donor allocation policy is often influenced by policy concerns beyond poverty reduction (e.g., global public goods) and historical ties and obligations, such as former colonies. Thus, it is not realistic for *all* foreign aid to be guided by a set of poverty-efficient criteria due to the various guiding interests nor would it be fit-for-purpose due to the different funding goals.

on the evidence supporting a needs-based allocation approach in this chapter before proceeding to issues related to effectiveness and performance-based allocation in the next two chapters.

Building on the conceptual framework from the introductory chapter, I start by explaining why average income is the most relevant measure of country need. I then examine various measures of development progress to better understand the magnitude of the development challenge. I explore the patterns in consumption poverty and a range of multi-dimensional poverty outcomes with an emphasis on differences between income groups. Second, I investigate the resources that are available to finance development efforts with an emphasis on government revenues. After examining the extent to which domestic resources are currently mobilized, I conduct a test comparing the poverty gap to the domestic resources potentially available. This produces the marginal tax rate (MTR) required for a country to hypothetically finance its own poverty reduction, which illustrates whether self-financing is feasible. Finally, I conclude with policy implications for donors allocating foreign aid, including an approximate level of average income at which aid should no longer be provided.

Conceptually, a donor focused on poverty reduction through economic growth should consider two allocation factors: need and effectiveness. Country need can be broken down into the magnitude of the development challenge and the domestic resources available. There are many possible measures of the development challenge, but average income is the best proxy for broad development progress. This is disputed, however, as "donors increasingly agree that an exclusive focus on income masks major underlying development challenges and is therefore inadequate for measuring multidimensional poverty" (OECD 2013). Thus, critics of average income have called for other measures, such as median income (Birdsall and Meyer 2015), SDG financing gaps (Manuel *et al* 2018), and many others. However, I show throughout this chapter that average income remains the single best measure of need – primarily because it is well correlated with the other suggested concepts but also because it conveniently describes both the development challenge *and* the resources available. In other words, average income provides both a good measure of a country's level of economic development and the amount of resources in an economy (scaled by population size) that could potentially be used by a country to finance its own development. These are precisely the two concepts that policymakers need to understand to make an informed decision about which countries have the greatest development need.

To help convince skeptics that average income is indeed the best measure of development need, I produce a set of data-driven observations describing country need across income groups. I establish that the highest rates of absolute poverty exist in the poorest countries in terms of average income. There is a clear trend across income groups when it comes to $2.15 per day poverty: LICs have the highest poverty rates by far; lower middle-income countries (LMICs) are more mixed with some countries that have nearly eliminated extreme poverty and other countries with extremely high poverty rates; and UMICs have very little extreme poverty left, except in a small handful of countries and even then, at relatively low rates. In terms of total numbers, the five countries with the greatest number of people living in poverty are India (129.5 million), Nigeria (62.2), Democratic Republic of the Congo (54.1), Tanzania (25.0), and Madagascar (21.3) – all LICs or LMICs. There is a less clear relationship between average income and the total number of people living in poverty because much depends on population size, but there are no UMICs among the ten countries with greatest number of people living in extreme poverty.

In addition to consumption poverty, I examine five different components of multi-dimensional poverty and their relationship to average income: energy, water, education, life expectancy, and child mortality. Like extreme poverty, I find that there is a clear separation between countries across income groups. At average incomes associated with richer UMICs, these deprivations become increasingly uncommon as long-run economic growth translates into eliminating multi-dimensional poverty in almost all cases. That is, economic growth is both necessary and sufficient for eliminating multi-dimensional poverty. I show that both consumption poverty and multi-dimensional poverty are strongly associated with average income, and there are significant differences between income groups. Though there is some overlap between the income groups, particularly with relatively poor UMICs that are outliers, LICs are undeniably at much lower levels of development when it comes to various measures of consumption poverty and multi-dimensional poverty. This makes a strong case that LICs in particular – but also LMICs – are in much greater need of development investments to better deliver basic public services, such as healthcare, education, and infrastructure. This is only one piece of country need, however, because resource availability factors into whether external assistance might be necessary to fill resource gaps.

Donors should also consider the domestic resources available to a country to finance development investments, as aid should be targeted to countries that are unable to mobilize sufficient resources to achieve their goals related to economic development. Because domestic resources are the most critical in terms of volume and relevance to aid allocation, I focus my analysis mainly on government revenues. I examine domestic resource mobilization in terms of both government revenues per capita and as a percentage of GDP. Revenues per capita are extremely low in LICs, but they get progressively larger as countries reach higher levels of average income, though the rate of increase slows for UMICs. In general, LICs are not able to generate significant revenues through taxation given their small economies, and they also have lower tax rates, on average.

I examine countries' tax potential relative to the magnitude of the poverty challenge by comparing taxable domestic resources to the extreme poverty gap. Following Ravallion (2010), I start by considering a taxable resource base that excludes the “middle class” between the extreme poverty line and a consumption line of $15 per day – thereby only taxing the “rich”. I show that the implied rates of taxation are excessively high for LICs and most LMICs but easily manageable for most UMICs to close the $2.15 per day poverty gap. The capacity to redistribute resources to close the extreme poverty gap is strikingly well aligned to the analytical income group classifications. I test the robustness of this finding by comparing the poverty gap to the resources of the non-poor and the entire economy (and not just the rich) and using higher poverty lines for LMICs and UMICs. When considering the potential to tax the non-poor’s resources above the poverty line, the capacity for LICs to redistribute improves dramatically – fewer than one third of LICs fall into the “low” capacity to redistribute. This is an extraordinarily low bar, however, and it is questionable policy to tax those just above the $2.15 poverty line to finance transfers to those just below that poverty line. LMICs also have a greater capacity to redistribute if taxing all of the non-poor even at the higher $3.65 per day poverty line. Finally, UMICs have a high capacity to redistribute regardless of the resource base even at a much higher poverty line.

A natural extension of my analysis is to identify at what level of average income most (or all) countries no longer require assistance.[12] While several UMICs have persistent extreme poverty, these countries are at the bottom of the UMIC group in terms of average income – only two countries above the IBRD line (of $7,065 in 2019) have an extreme poverty headcount ratio above three percent. For multi-dimensional poverty, a number of UMICs below the IBRD line still experience significant multi-

[12] There are several reasons why a threshold for cutting off assistance is more desirable that a continuous function that phases out aid as a country reaches higher levels of income. First, there are fixed costs associated with donors having an in-country presence, so it doesn’t make sense to allocate assistance below that (non-zero) amount. Second, donors are notoriously undisciplined at exiting (d’Orey *et al* 2019), so a threshold imposes rules on the process. In practice, a cut-off threshold looks more like a tapering of assistance anyway, as the final allocations are spent over time. Finally, a continuous function would force a decision as to where the allocation formulate zeroes out at some point in the distribution anyway – whether at an arbitrary threshold or at the UMIC/HIC threshold (when transfers are no longer counted as ODA). Absent such a threshold, many donors, especially those with a 0.7% ODA/GNI commitment, default to the UMIC/HIC line, and that is arguably far too high. These issues will be explored further in the concluding chapter.

dimensional poverty, while almost all countries above the IBRD line have eliminated multi-dimensional poverty. This suggests that approximately the IBRD line may be a reasonable expectation for countries to have eliminated both consumption and multi-dimensional poverty. In terms of domestic resources, there are no UMICs with a low capacity to redistribute and only one UMIC with a medium capacity to redistribute. All countries above the IBRD line should be able to easily finance their own poverty reduction, as the MTR required is less than one percent of the resources of the richest segment of their population. This all strongly suggests that a threshold for cutting off assistance should not be the IDA threshold employed by the World Bank or the LMIC/UMIC threshold used by MCC, but rather that the IBRD line may be the most appropriate threshold for cutting off grant assistance entirely.

I make two contributions to the needs-based allocation literature in this chapter – the first is clarifying the best measures of need and the second is identifying a threshold for cutting off aid. The first contribution is that I contest the increasingly popular narrative that average income is an insufficient measure of development need. Kappel (2017) argues that average income is “neither a good approximation to prevalent poverty levels nor to the capacity to fight poverty with own resources,” and I show this assertion to be false on both counts. By examining each of these components of development need in detail, I find that countries at lower levels of average income experience the greatest development challenges and have the fewest resources to address those challenges. These findings constitute a compelling argument that foreign aid should be allocated to the poorest countries in terms of average income. My second contribution is an evidence-based policy recommendation of where to implement a cut-off for assistance in terms of average income. Across both components of country need – the magnitude of the development challenge and the resources available – I identify a common threshold around the IBRD line, which was $7,065 in 2019. Above this line, most countries have eradicated consumption and multi-dimensional poverty, and they have sufficient domestic resources to address the remaining problem to the extent it still exists.

The policy implications of my analysis are that LICs and LMICs should be strongly favored in aid allocation models and that aid should be cut off above the IBRD line. LICs and LMICs are in much greater need of external financing, as the magnitude of the challenge is much greater across all measures of development outcomes. Appeals to continue providing aid to UMICs often appeal to multi-dimensional poverty, but even if a policymaker wants to expand their focus to include multi-dimensional poverty, the evidence leads to the same implication – a focus on LICs and LMICs. In addition, domestic revenues are extremely low in LICs, and countries are increasingly able to finance their own development as they reach higher levels of average income, particularly UMICs. The capacity to redistribute resources to close a country's poverty gap also increases with average income, as the poverty gap is much larger in LICs and LMICs while domestic resources are much larger in UMICs. Putting it all together, the two different components of country need both point in the same direction – with the same policy implication of a strong focus on allocating assistance towards the poorest countries.

While some relatively poor UMICs may have some persistent poverty that might warrant donor attention, the IBRD line provides an evidence-based option as a threshold for cutting off assistance. I find that a reasonable cut-off for transitioning away from aid is around $6,000 to $7,000 GNI per capita – roughly the IBRD line – when considering whether countries can be expected to eliminate poverty and/or self-finance their own poverty eradication. This is consistent with the strategic vision of many donors to transition countries away from grant assistance towards financing their own development. While this is an intuitive finding, it is often not implemented in terms of aid allocations (Dassinayake *et al* 2020). Put simply, most grant financing should go to the very poorest countries and should be cut off around the IBRD line in terms of average income. At the very least, grants going to UMICs should be approached differently and have large impacts in addressing outstanding development challenges, while starting to transition the diplomatic engagement away from an aid relationship.

The development challenge

I must first define "need" and why it is an important consideration for aid allocation. Crosswell (1980) asked the definitional question and explains that "…need for outside assistance depends on the extent of the 'problem' to be addressed and domestic resources available to direct towards the problem." Thus, country need can be broken down into two components: the magnitude of the development challenge and the resources available (Crosswell 2015). The magnitude of the development challenge can be measured in numerous ways – from aggregate, economy-wide resources to distributional measures to multi-dimensional poverty, though I explain below why average income is the single best measure of the development challenge. I set aside the effectiveness criterion in this chapter to focus on need, but I examine the evidence for a performance-based allocation system in the next two chapters. I examine resource availability in the next section as a complement to this analysis.

Beyond the obvious point that taxpayers should not have to finance assistance to countries that do not need it, Kenny (2020) makes several arguments in support of allocating assistance to countries where the development challenge is the greatest. He argues that (1.) the marginal utility of income for those living under the extreme poverty line would be higher, including greater (proportional) spending on absolute necessities (Engels curve); (2.) there is evidence on subjective life satisfaction that suggests that there is a stronger relationship between income growth and life satisfaction at lower levels of income; (3.) there are diminishing returns to health (Preston curve) in that increases in income improve health outcomes greatly at lowers levels of income, but this relationship tapers off at higher levels of income;

and (4.) the nearly universal prevalence of progressive tax schemes and social safety nets suggests that reducing poverty and inequality is a widely held public policy goal. Brainard (2003) reinforces this thinking: “In principle, pure development assistance should be allocated to the investments with the highest marginal value, determined by the extent of need (or the marginal social value)…”

In this section I examine various measures of development progress to better understand the magnitude of the development challenge in developing countries. I start by explaining why average income is the most relevant consideration for country need. I then explore the relationship between various measures of poverty and average income, with a particular focus on comparisons between income groups. I examine descriptive statistics and patterns in both consumption poverty and a range of multi-dimensional poverty outcomes to determine whether average income is indeed a useful measure of the development challenge and whether there are differences between the income groups. I conclude by identifying a threshold for cutting off assistance based on the preceding analysis.

Average income

McKee *et al* (2020) conclude that “There are many possible proxy measures for country need… the most straightforward of which is per capita income.” Though it has flaws,[13] average income is widely considered the best proxy for broad development progress (Crosswell 2015), and this is for at least two reasons. First, it represents the average level of income across the entire economy. While average income is sometimes criticized for telling us nothing about the underlying distribution of income in a particular country (Birdsall and Meyer 2015), it tells us more about global inequality than is appreciated. Though within-country inequality is more important when it comes to global inequality, between-

[13] For example, standard measures of average income, such as GNI per capita, do not capture work in the home or environmental damage, among other economic activity and externalities.

country inequality comprises roughly *one third* of global inequality, though this proportion has been decreasing (Chancel *et al* 2022). This is important to keep in mind when considering alternative broad measures of development progress; for instance, median income tells us something about the distribution of income, but it only gives one point in the distribution – the 50^{th} percentile – and gives only a rough sense of the total resources in the economy.

Second, average income provides an additional, extremely valuable piece of information when considering country need. It is a measure of the total resources in an economy scaled by the number of people. That is, if you multiply GDP per capita times the total population, you'll get total GDP, and total GDP is the sum of economic activity in a given country that could potentially be taxed and redistributed. This is extremely useful in considering country need as it relates to resource availability, because it measures the *potential* for redistributing resources, while remaining agnostic to the actual distribution of income, which is at least partially a result of political choices and public policy, and the rate of taxation or level of government spending. In other words, average income provides an apples-to-apples comparison of the resources available in a country that could potentially be re-distributed.

Therefore, average income not only gives us a good measure of the level of a country's economic development, but it also tells us the amount of resources in an economy that could potentially be used by a country to finance its own development. As discussed above, these are precisely the two concepts we need to understand to make an informed decision about which countries have the greatest development need. Granted, there are critical measures of the development challenge beyond average income and there are measures of a country's resources (and external resources) that extend beyond just average income. However, no measure tells us as much in one single number that is comparable across countries. Thus, it is a powerful instrument for policymakers when allocating assistance.

If one accepts that average income is a useful measure of country need – and perhaps the best available to us – then there comes the question of country groupings for the sake of analysis. While the World Bank's income classification groupings are admittedly arbitrary and without underlying economic meaning (Kenny 2014), they are nonetheless useful heuristics for referencing a country's level of development and grouping countries at a similar level of development together for the purposes of analysis (hence why they are technically named the "analytical classifications" by the World Bank). As a result, I will make extensive use of these analytical income classifications for the purposes of drawing comparisons between countries at different levels of development.[14] I will show that there are meaningful differences between these income groups, particularly between LICs and LMICs versus UMICs. While these distinctions may blur at the threshold (Kenny 2014), the median LMIC and especially the median LIC are at a dramatically lower level of development than the median UMIC across a range of diverse development outcomes and measures.

Extreme poverty

Starting with consumption poverty, the $2.15 per day international poverty line – previously "dollar-a-day", then $1.25, then $1.90 – is the most commonly cited poverty line. Consumption below this threshold is often referred to as "extreme poverty" given the extremely low standard of living it implies. While it is arguably penuriously low (Pritchett 2006), it is close to the bare level of consumption a person can survive on and therefore a useful low-bar threshold. Furthermore, ending absolute poverty (based on the $2.15 per day poverty line) is the foremost goal as articulated by the international community as

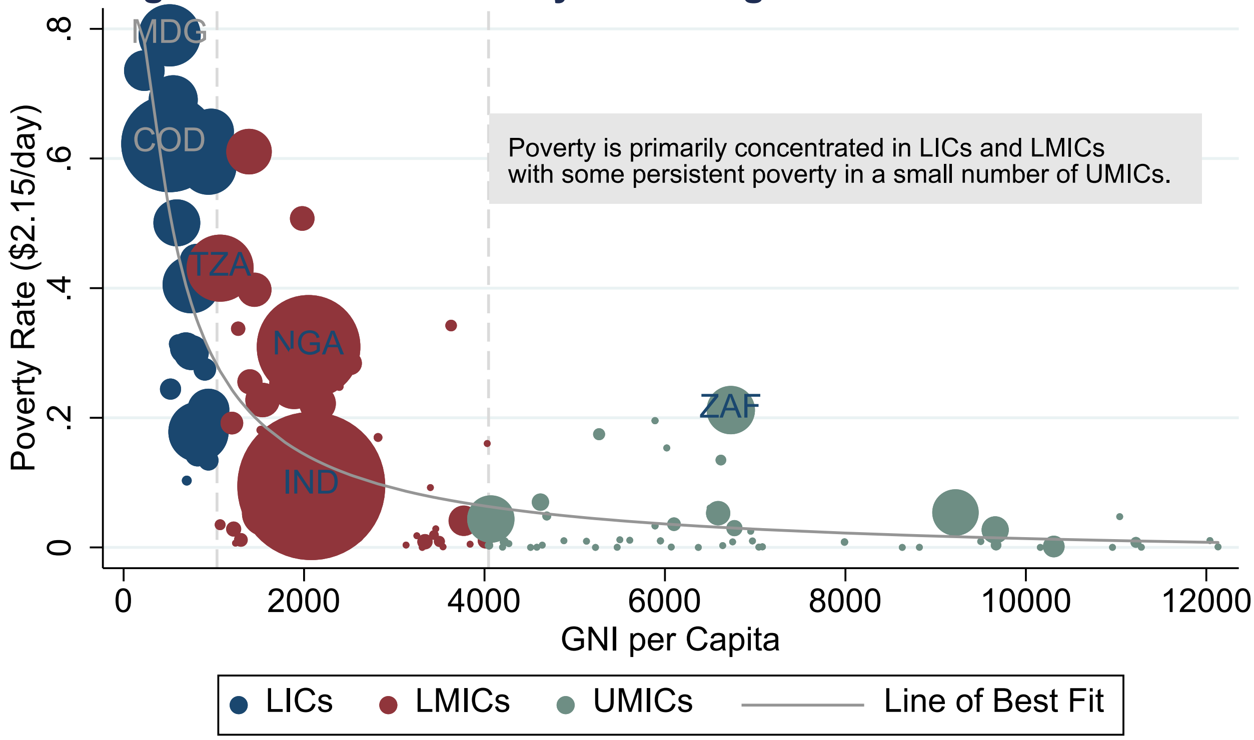
[14] I do so with the caveat that there are often greater differences in income within income groups than there are between them. For instance, the average income between an LMIC and a UMIC might be less than $100 per year, whereas the difference between a country at the bottom of the UMIC grouping versus a country at the top of the group might be separated by close to $8,000 (as the UMIC group extended from $4,046 up to $12,535 in 2019).

enshrined in Sustainable Development Goal 1 (and Millennium Development Goal 1 before that). While development objectives should certainly aim higher than this “extreme” poverty line, it is hard to argue that anyone should have to live on less, so it is a good place to start when considering the allocation of development resources. That is, countries with a high proportion of their populations living on less than $2.15 per day face a massive development challenge. Even a small amount of resources can go a long way in improving these people’s lives if it is spent effectively to raise their living standard.

If one accepts the argument that $2.15 per day poverty is a reasonable measure for the goal of ending poverty, then the next question becomes: Where is the highest prevalence of extreme poverty? This can be considered across many dimensions – for instance, sub-nationally, demographically (gender, age, ethnicity, urban vs. rural, etc.), etc. From the donor perspective, however, the policy maker is most concerned with the country-level poverty rate for the purposes of allocating assistance, at least when considering high-level trade-offs in terms of where best to spend scarce grant resources, particularly at a government-to-government level (or multilateral-to-government in the case of the World Bank). This is because ODA is typically allocated at the country level, and the headcount ratio (or poverty rate) is the most direct measure of the prevalence of poverty in a given country. Furthermore, while there are many ways to answer where poverty exists using the $2.15 poverty line, the most easily understood is either in terms of the countries with the greatest total number of people in poverty or the countries with highest poverty rate.[15] When considering the allocation of assistance, a scaling factor for population can be used, so I’m most interested in the poverty rate, or headcount ratio, and not the total number of poor, though I will also consider total numbers as part of my analysis.

[15] For instance, the Foster-Green-Thorbecke poverty measurements includes additional measures of poverty, such as the poverty gap and severity of poverty indices. The headcount ratio and the total number of poor remain the easiest to understand and interpret, however. As a result, they are the most commonly cited poverty measures.

Figure 1. Extreme Poverty vs Average Income



Note: Poverty rate at $2.15 per day plotted against its GNI per capita (Atlas) in 2019.
The bubbles are weighted by the size of the population living in poverty.
Source: World Bank Poverty and Inequality Platform, World Develoment Indicators.

It is undeniable that most absolute poverty at lower poverty lines exists in poorer countries.[16] Figure 1 shows the $2.15 poverty rate plotted against average income in GNI per capita terms for 2019. This is not a surprising pattern as economic growth is well correlated with poverty reduction – higher average incomes are a result of long-term economic growth – but there is a striking separation across income groups. Many of the LICs in blue have extremely high levels of extreme poverty above 60 percent, whereas very few LMICs come close to these high levels of poverty. No LICs had a poverty rate below 10 percent, and the median (mean) poverty rate was 42.4 (43.8) percent. LMICs are more heterogeneous in

[16] It is necessary to draw a distinction between absolute and relative poverty. Absolute poverty refers to measures of poverty as they relate to a common poverty line in nominal terms, such as the international poverty line of $2.15 per day. While imperfect, a common poverty line allows researchers and policymakers to make cross-country comparisons. On the other hand, relative poverty refers to measures of poverty that sets a poverty line based on the distribution of income or consumption within a given context. Relative poverty lines are usually established by the national government and change over time, e.g., a common rule is to set the national (relative) poverty line at half of median income. As a result, this approach does not lend itself to cross-country comparisons.

their prevalence of extreme poverty, from seven countries with less than 1 percent to two countries above 50 percent (Republic of the Congo and Zambia). The mean (mean) poverty rate in LMICs is 9.2 (14.9) percent – less than one quarter of the median LIC. UMICs have an almost uniformly low extreme poverty rate with the median (mean) being just 0.9 (3.0) percent – more than seven in ten (34 of 48) UMICs with data have an extreme poverty rate below three percent. Only one UMIC has a poverty rate above 20 percent (South Africa) with an additional three countries above 15 percent (Suriname, Namibia, and Belize). This suggests that average income is far more important in predicting poverty than inequality, as South Africa is the one of the very few exceptions with its extremely high level of inequality that keeps poverty rates high despite its UMIC status. There is a clear trend across income groups when it comes to $2.15/day poverty – LICs have the highest poverty rates by far, LMICs are more mixed (with some countries that have nearly eliminated extreme poverty and other countries with extremely high extreme poverty rates), and UMICs have very little extreme poverty left, except in a small handful of countries and even then at relatively low rates compared to almost all LICs.

The size of the markers in Figure 1 are weighted by the number of poor in that country. The five countries with the greatest number of people living in poverty are labeled: India (129.5 million), Nigeria (62.2), Democratic Republic of the Congo (54.1), Tanzania (25.0), and Madagascar (21.3). Of course, a country must not only have a high poverty rate to have a large number of people living in poverty, but it must also have a relatively large total population. As a result, there is a less clear relationship between average income and the number of people living in poverty. Regardless, due to their lower poverty rates overall, there are no UMICs among the countries with the top ten number of people living in poverty. South Africa (12.4 million) and Indonesia (11.9) are the UMICs with the largest number of people living on less than $2.15 per day. See Appendix 1 for a side-by-side comparison of the countries with the highest poverty rates and total number of people living in poverty – there is some overlap, but both lists are mostly populated by LICs.

A natural extension of this analysis is to identify at what average income most (or all) countries have eliminated extreme poverty.[17] While several UMICs (above $4,045 in 2019) have persistent extreme poverty above ten percent, these countries are at the bottom of UMIC category in terms of average income: Namibia at $5,270 GNI per capita (17.5 percent); Belize at $5,890 (19.6), Suriname at $6,020 (15.3), Botswana at $6,620 (13.5), and South Africa at $6,730 (21.2). Only two countries above the IBRD line (of $7,065 in 2019) have an extreme poverty headcount ratio above three percent: Brazil (5.4 percent) and St. Lucia (4.7 percent). This suggests that approximately the IBRD line may be a reasonable expectation for countries to have eliminated extreme poverty, though it is unlikely South Africa will completely eliminate poverty as it crosses that line, so this could be a somewhat conservative threshold.

Multi-dimensional poverty

If one does *not* accept that $2.15 per day poverty is an adequate measure of deprivation, then I might instead ask: "Where is the highest prevalence of multi-dimensional poverty?" While multi-dimensional poverty indices like the UN's Human Development Index (HDI) or Oxford's Multi-Dimensional Poverty Index (MPI) are useful for aggregating multiple social outcome indicators, they are difficult to interpret as they are inherently combining information from multiple indicators into one number. Ravallion (2011) criticizes multi-dimensional indexes by asking what purpose they serve (that is not achieved by a consumption measure), observing that the choice of weights is inherently arbitrary and unnecessary,

[17] I use the World Bank target of three percent as a proxy for "ending extreme poverty".

and that analysts should therefore examine multi-dimensional outcomes individually. In other words, it is better to have a credible set of indicators than one multi-dimensional index.[18]

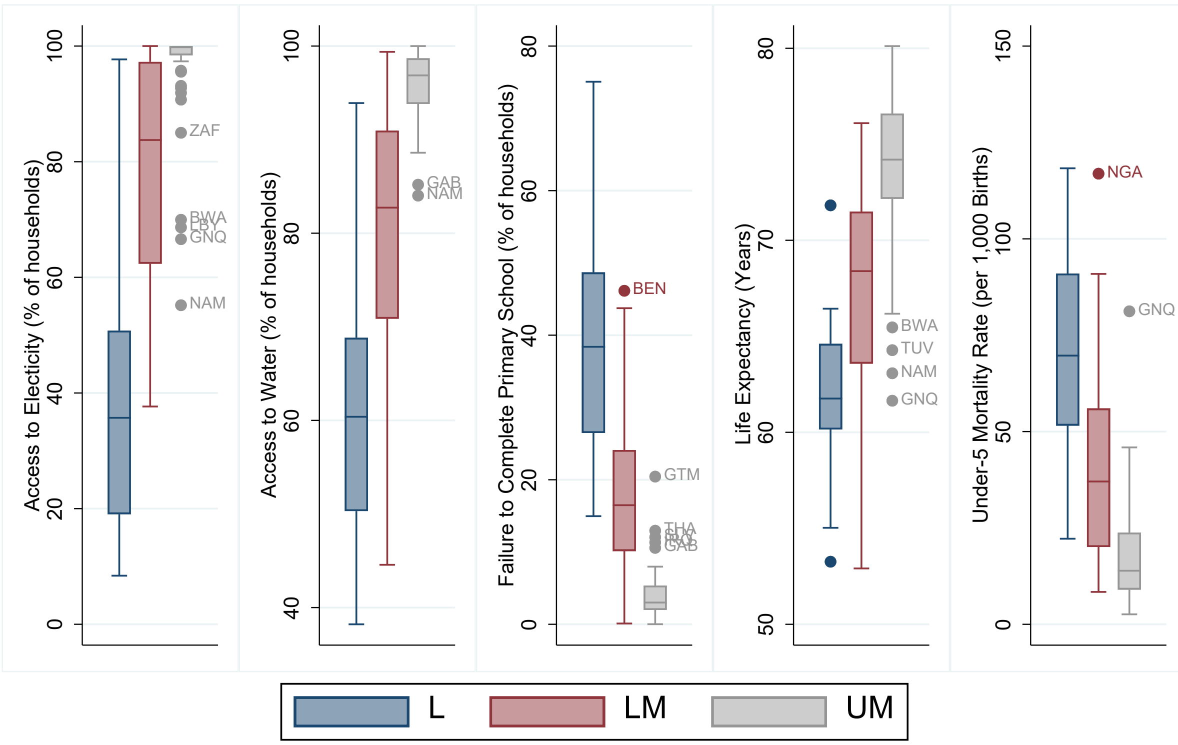
With this in mind, I examine five different social outcomes as a proxy for multi-dimensional deprivation and observe the differences between income groups. The UN's HDI equally weights living a long and healthy life (life expectancy), becoming knowledgeable (expected year of schooling and mean years of schooling), and having a decent standard of living (as measured by GNI per capita). I use life expectancy as one of my indicators, like HDI, but I use the educational indicator from MPI, primary school completion. Oxford's MPI also has three dimensions – health, education, and living standards – but it utilizes ten different indicators. In addition to life expectancy and primary schooling, I examine under-five mortality, access to water, and access to electricity. Between the five indicators, I cover all three dimensions of both HDI and MPI.

The box and whiskers chart in Figure 2 shows the distribution of social outcomes across income groups in 2019 for life expectancy, under-five child mortality, primary schooling, access to water, and access to electricity. First, it is critical to note that these outcomes are well correlated with each other and, importantly, with average income. I show these partial correlations in Appendix 2. The boxes in Figure 2 show the inter-quartile range – the line through the middle is the median, while either end of the box is the 25th and 75th percentile for that particular income group. The whiskers show the minimum and maximum value where there are no outliers. Where there are outliers identified, the whiskers are 1.5 times the inter-quartile range (i.e., between the 25th and 75th percentile or the length of the box). Like

[18] I must credit a former USAID colleague and economist, Don Sillers, for making this point simply and powerfully. He observed that you can combine your inseam and waist to get a pant size index, but the pants are unlikely to fit.

extreme poverty, there is a clear separation in development outcomes between income groups, as evidenced by the limited overlap in the inter-quartile ranges from one income group to the next.

Figure 2. Multi-Dimensional Poverty by Income Groups



Note: Various development outcomes across income groups in 2019.
Source: Multi-Dimensional Poverty Index, World Develoment Indicators.

For electricity, the median (mean) proportion of the population with access to electricity is just 40.4 (38.7) percent for LICs, while it is 83.8 (77.6) percent for LMICs and nearly universal at 99.8 (96.5) percent for UMICs. The 25th percentile for LICs is below one fifth (19.0 percent), and the minimum extends down below ten percent (6.7 percent in South Sudan). There is significant variation among LMICs with the inter-quartile range extending from nearly 100 percent (97.3 percent) at the 75th percentile down below two thirds (62.4 percent) at the 25th percentile with the minimum all the way below 40 percent (37.7 percent in Tanzania, which is also the poorest LMIC in terms of GNI per capita). There are more outliers for UMICs for electricity than the other indicators shown here with four outliers ranging from 55 to 70 percent (Botswana, Libya, Equatorial Guinea, and Namibia from highest to lowest), which is well below the inter-quartile range of 98.6 to 100 percent. The inter-quartile range is very narrow because the vast majority of countries in the UMIC income group have near-universal access to electricity – all but five UMICs had greater than 90 percent access, and more than three quarters (35 of 46) have 99 percent access to electricity or greater.

A similar trend is found with access to water, as measured by the proportion of the population with access to an improved drinking water source, such as piped water to households or dug wells. The inter-quartile ranges for each income group line up almost perfectly end to end with each other. One way to interpret this is that the bottom three quarters of LICs have a lower rate of access to water than the top three quarters of LMICs, and the same is true between LMICs and UMICs. The median (mean) LIC has 60.4 (61.5) percent of their population with access to water, whereas the median LMIC has about 82.7 (80.7) percent access, and the median UMIC has 96.9 (95.9) percent. There are only two outlier UMICs that are similar to the median level of access to water for LMICs (Namibia and Gabon at 84.0 and 85.2 percent, respectively), and nearly two thirds of UMICs (36 of the 55 with data) have near-universal access to water of 95 percent or greater with only three below 90 percent (including the Marshall Islands). This again shows that there is significant separation between the income groups.

Like water, there is significant separation between income groups for education with no overlap between the inter-quartile ranges for each income group. My education indicator is from MPI, which is measured as the percentage of households with at least one member having not completed six years of schooling (that is old enough to have done so). The median (mean) LIC has 39.8 (39.6) percent of households experiencing this deprivation, whereas the median (mean) LMIC has nearly half that proportion with 24.1 (18.0) percent. Not completing primary school is relatively rare in UMICs, as the median (mean) UMICs has 3.0 (4.3) percent of its households with this deprivation. Of the 41 UMICs with data for this indicator, more than two thirds (28 of 41) have more than 95 percent of households in which all members have completed primary school. Only five have more than ten percent of households experiencing this deprivation, and only one UMIC is above 20 percent (Guatemala). This is in contrast to LICs in which only three countries (of the 22 with data) have fewer than 20 percent of households that are deprived (The Gambia, DRC, and Yemen).

There is relatively more parity between LICs and LMICs for life expectancy – but not UMICs. The average life expectancy at birth for the median (mean) LIC is about 62.4 (62.5) years, whereas it is about 68.4 (67.7) years for the median LMIC and 74.1 (73.6) years for the median UMIC. While the median countries are fairly far apart between income groups, the country with the minimum life expectancy for both LICs and LMICs is similar at about 53 years. UMICs have much higher life expectancy, however, with only three countries below 65 (Equatorial Guinea, Namibia, and Tuvalu), and 49 of the 58 UMICs with data above 70 years of age on average. Like with the previous outcomes examined, the top three quarters of UMICs are above the bottom three quarters of LMICs with only seven UMICs below the LMIC median of 68.4 years. While there has been significant convergence in life expectancy largely due to technological advancements (Kenny 2011), there are still significant differences across income groups, particularly when countries reach the level of average income of the median UMIC – indeed, the 16 UMICs with the lowest life expectancy are all below $7,000 GNI per capita.

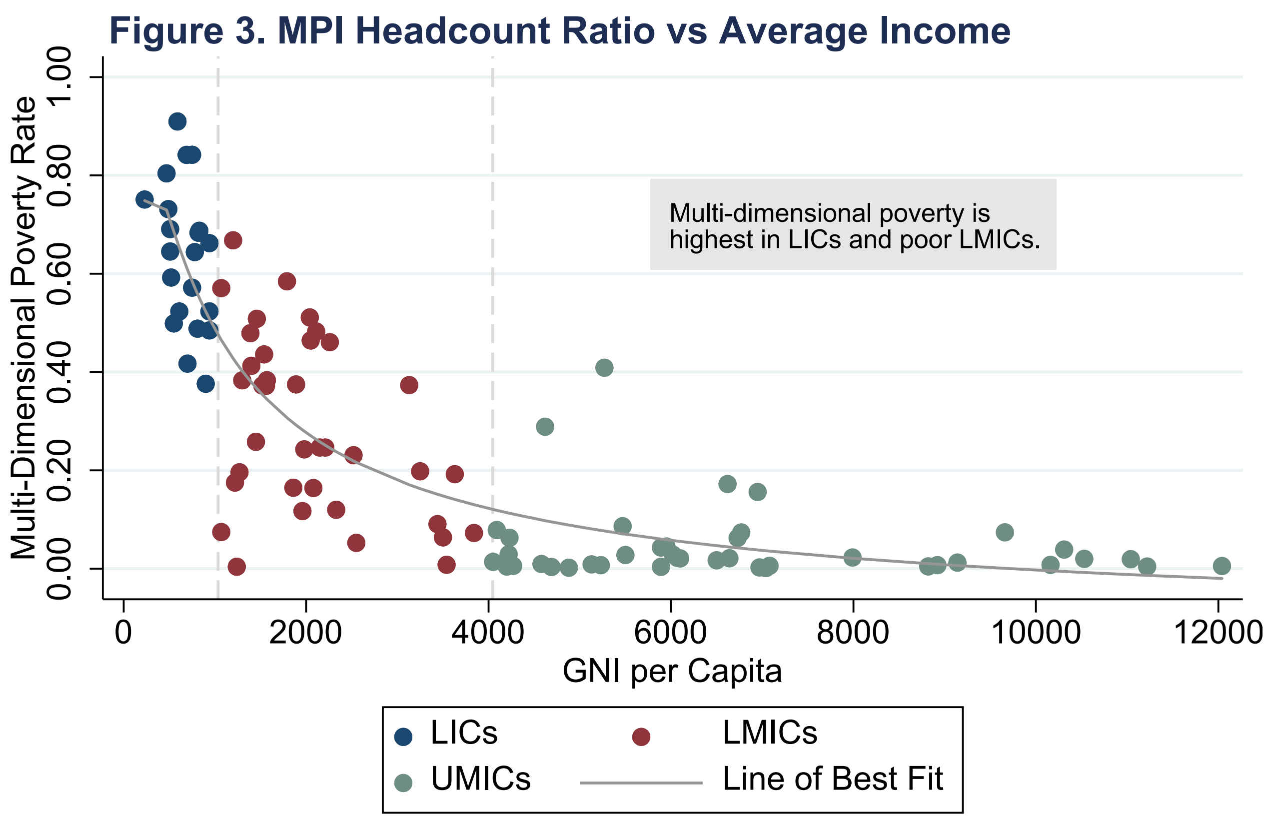
Finally, under-five child mortality shows less stark differences between income groups than the other indicators. Nonetheless, LICs perform significantly worse in terms of child mortality with the median (mean) country having an infant mortality rate of approximately 66.5 (70.0) deaths per 1,000 live births. That is more than three quarters greater than the median LMIC at 37.1 (41.7) and nearly five times the median UMIC at 14.1 (18.2). There is not as much divergence between LICs and LMICs, as the inter-quartile range overlaps between LICs and LMICs, and the LMIC (Nigeria) with the highest under-five mortality rate (116.9) is almost as the same level as the LIC (Somalia) with the highest rate (118.3) overall, so progress on under-five mortality appears to lag when moving from LIC to LMIC, at least compared to the other social outcomes examined above. However, very few UMICs have a high infant mortality rate with nearly four fifths of UMICs (47 of 59) below 25 deaths per 1,000 live births and only five countries above the LMIC median.

These trends in the multi-dimensional poverty data suggest that rising average incomes translate into gains in other development outcomes in almost all cases. All of the deprivations I have examined are increasingly uncommon at higher levels of average income, particularly at the high end of the UMIC income group. This observation is consistent with the findings of Pritchett and Lewis (2022) that "economic growth is enough and only economic growth is enough." They claim that for *any* measure of basic human progress, such as the multi-dimensional poverty outcomes presented above, there will be a strong, non-linear relationship to average income. The relationship is non-linear because there is a stronger elasticity at lower levels of income, i.e., countries improve more rapidly on basic needs as they raise their average incomes from very low levels.

This non-linear relationship is apparent in the box and whiskers charts in Figure 2, but it is even more obvious in the scatter plot presented in Figure 3, which shows the MPI's multi-dimensional poverty headcount ratio (not the index) plotted against average income.[19] As shown by the line of best fit, there are rapid reductions in multi-dimensional poverty at low levels of average income as countries

[19] MPI considers a person to be multi-dimensionally poor if they are deprived in at least one third of the weighted MPI indicators, which is used to obtain a headcount ratio (H) as a proportion of the total population. The index ratio of MPI also includes a measure of the intensity of the deprivation (A) to produce the adjusted poverty headcount ratio (M_0) by multiplying H and A (M_0 = H x A). I use the MPI headcount ratio as it is more analogous to the $2.15 per day headcount ratio used above.

experience economic growth and move from the LIC income group (blue dots) to the LMIC income group (red dots). The relationship starts to flatten out at levels of development associated with richer LMICs and then there are limited improvements once a country reaches UMIC status. This holds true across all of the indicators examined here, and Pritchett and Lewis (2022) show that this holds across all measures of basic human needs.



Note: Multi-dimensional poverty headcount ratio plotted against GNI per capita (Atlas) in 2019.
Source: Oxford Multi-Dimensional Poverty Index, World Develoment Indicators.

Pritchett and Lewis (2022) further claim that a high level of average income is both necessary and sufficient to achieve progress on these social outcomes, as no country has achieved these outcomes at low levels of income (necessary) and no country has failed to achieve these outcomes at low levels of income (sufficient). They make this point by comparing the social outcome indicators of high-income countries (HICs) to that of developing countries. However, I only examine developing countries, so this trend is not as clear as there is not universal attainment of these basic human needs once a country reaches the UMIC income group like there is with the HIC group.[20] Nonetheless, across the indicators examined above, even the worst-performing UMICs that are outliers still have lower levels of deprivation than most LICs and even the best-performing LICs do not approach the median UMIC.

Figure 3 reinforces that economic growth is both necessary and sufficient. As with consumption poverty, there are a small number of outliers that are relatively poor UMICs, but by the time a country reaches about $7,000 GNI per capita, there is very little multi-dimensional poverty. Only two (of the 13) UMICs with a multi-dimensional poverty rate above three percent have a GNI per capita above $7,000 annually (China and Mexico). Once countries approach the threshold to become a HIC, there is very limited multi-dimensional poverty – that is, economic growth is sufficient to eliminate multi-dimensional poverty and achieve basic human needs. On the other hand, there are no LICs that have even come close to eliminating multi-dimensional poverty. Togo has the lowest multi-dimensional poverty at 37.6 percent, and this is only higher than one UMIC, Namibia at 40.9 percent. This suggests that economic growth is also necessary to significantly reduce multi-dimensional poverty and achieve basic human needs.

As in the previous section, I conclude by asking what level of average income a country might be expected to end multi-dimensional poverty. Again, a number of UMICs below the IBRD line experience significant multi-dimensional poverty as measured by the MPI headcount ratio: Guatemala at $4,620 GNI per capita (28.9 percent), Namibia at $5,270 (40.9 percent), Botswana at $6,620 (17.2 percent), and Gabon at $6,950 (15.6 percent) are all above ten percent. Above the IBRD threshold, only two countries (of the 14 with data) are above three percent: Mexico (7.4 percent) and China (3.9 percent). This

[20] Again, this is at least partially because the lower end of the UMIC income group (at above $4,045) is still a relatively low level of development compared to a HIC (starting at $12,535 in 2019).

suggests that almost all countries above the IBRD line have eliminated multi-dimensional poverty. Again, Botswana and Gabon are close to the IBRD threshold, and they both have significant levels of multi-dimensional poverty, so they are unlikely to completely eliminate multi-dimensional poverty as they cross the IBRD threshold. Nonetheless, the seven countries in terms of average income between Gabon ($6,950) and Mexico ($9,660) that have MPI data all have an MPI headcount ratio of 2.3 percent or lower, and five are below one percent, so Botswana and Gabon may be the exceptions to the rule.

In sum, I have shown that both extreme poverty and multi-dimensional poverty outcomes are strongly associated with average income, and there are significant differences in these outcomes across income groups. Though there is some overlap between the income groups, particularly with relatively poor UMICs that are outliers, LICs are undeniably at much lower levels of development when it comes to various measures of consumption poverty and multi-dimensional poverty. This makes a strong case that LICs in particular – but also LMICs – are in much greater need of development investments to better deliver basic public services, such as healthcare, education, and infrastructure. It also strongly suggests that a threshold for cutting off assistance should not be the IDA threshold ($1,945 in 2019) or the LMIC/UMIC threshold employed by MCC, but rather that the IBRD line may be the most appropriate threshold for cutting off grant assistance. The analysis above is only one piece of country need, however, so I will next explore the resources available across income groups to determine whether external grant assistance might be necessary to fill resource gaps.

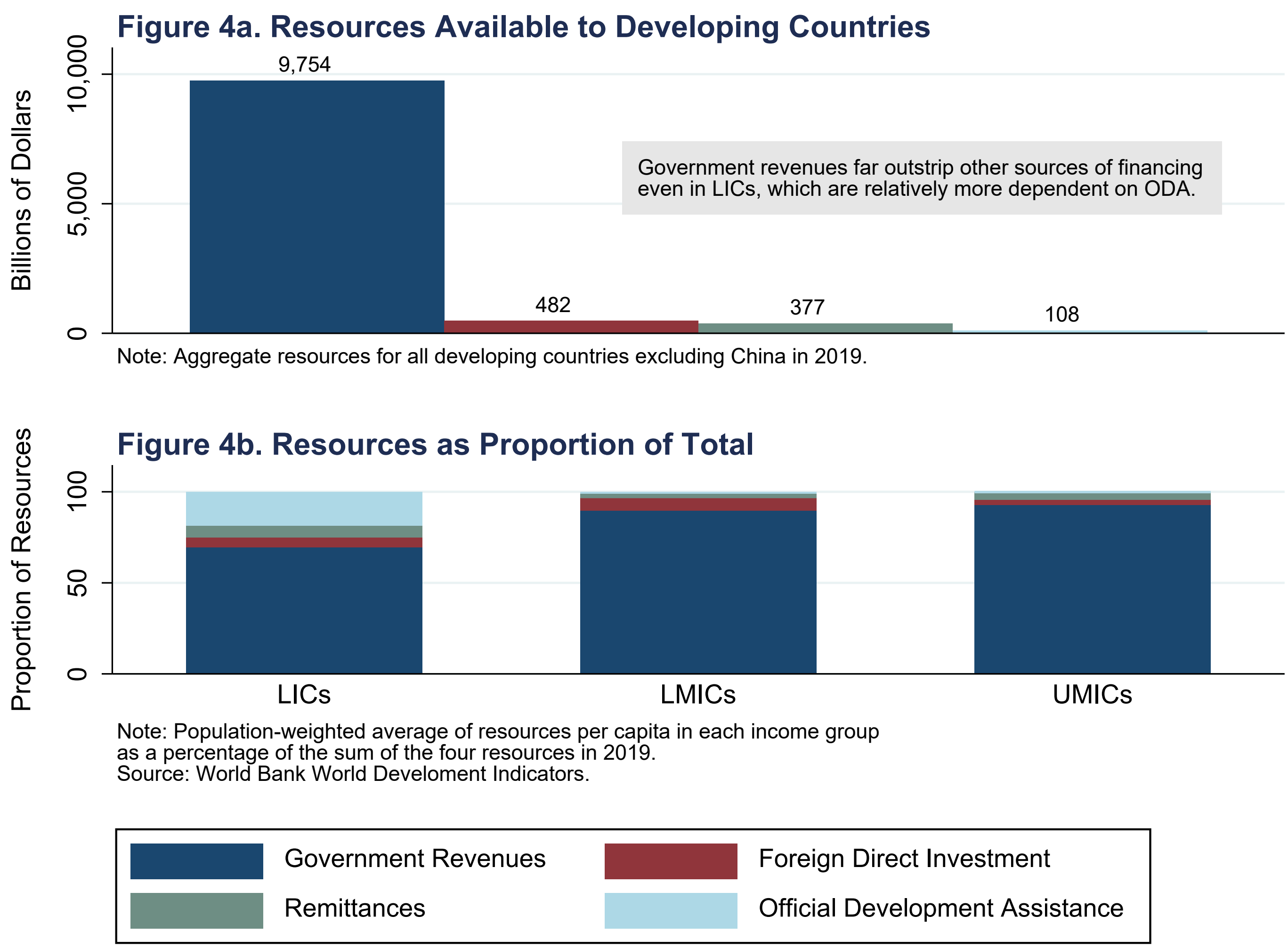
Resource availability

In addition to the magnitude of the development challenge, donors should consider the resources available to a country to finance its own development investments – whether infrastructure or social services or social protection. McKee *et al* (2020) succinctly summarized the thinking related to resource availability: “If therefore a primary objective of aid is poverty reduction, and a primary characteristic of ODA is its potential to provide concessional resources to countries that have difficulty in accessing other financing, ‘high quality’ aid means aid that is well-targeted to such contexts.” The most obvious place to start is the country’s tax base – that is, the amount of public resources at its disposal to appropriate to public spending needs. From the donor perspective, however, this is only part of the story as many low-income countries have a low tax administration capacity, which leads to low effective tax rates. Furthermore, tax policy is at least partially a political choice, so government revenues are partly reflective of political interests and the broader governance environment as opposed to solely just the resources available. That is, a country may have a low effective tax rate because they *choose* not to tax their elite or lack the administrative capacity to enforce their tax laws. Therefore, a measure of the total taxable resources available may be more appropriate. I explore this further later in this section.

There are many different sources of development finance, including domestic savings, foreign direct investment, remittances (both cross-border and domestic, such as urban-rural), government revenues, official flows, and ODA, among others. It is important to note that these other forms of development finance play a critical role; however, I will show that these resources, including ODA, pale in comparison to the domestic resources available to developing countries, except in the poorest countries – which in itself is a compelling reason for donors to direct their grant resources towards the poorest countries.

From a donor perspective, government revenues are arguably the most important source of development finance because they are much larger than all international resource flows to developing countries (Development Initiatives 2013) and because it is within the policy domain of the country partner government. Figure 4a shows various international resources to developing countries and confirms that domestic revenues outstrip FDI, remittances, and ODA by roughly an order of magnitude.

The figure includes resources to all developing countries except China, which would add more than $3.6 trillion in additional government revenues. Furthermore, Figure 4b shows that domestic revenues are the dominant resource in LMICs and especially UMICs, while only LICs have a significant proportion of their resources available from an international flow. This also shows the relative aid-dependence of LICs with ODA accounting for about 18.5 percent of resources per capita (population-weighted across countries in the income group to represent the proportion of the total resources per capita for the average person living in a country in that income group).



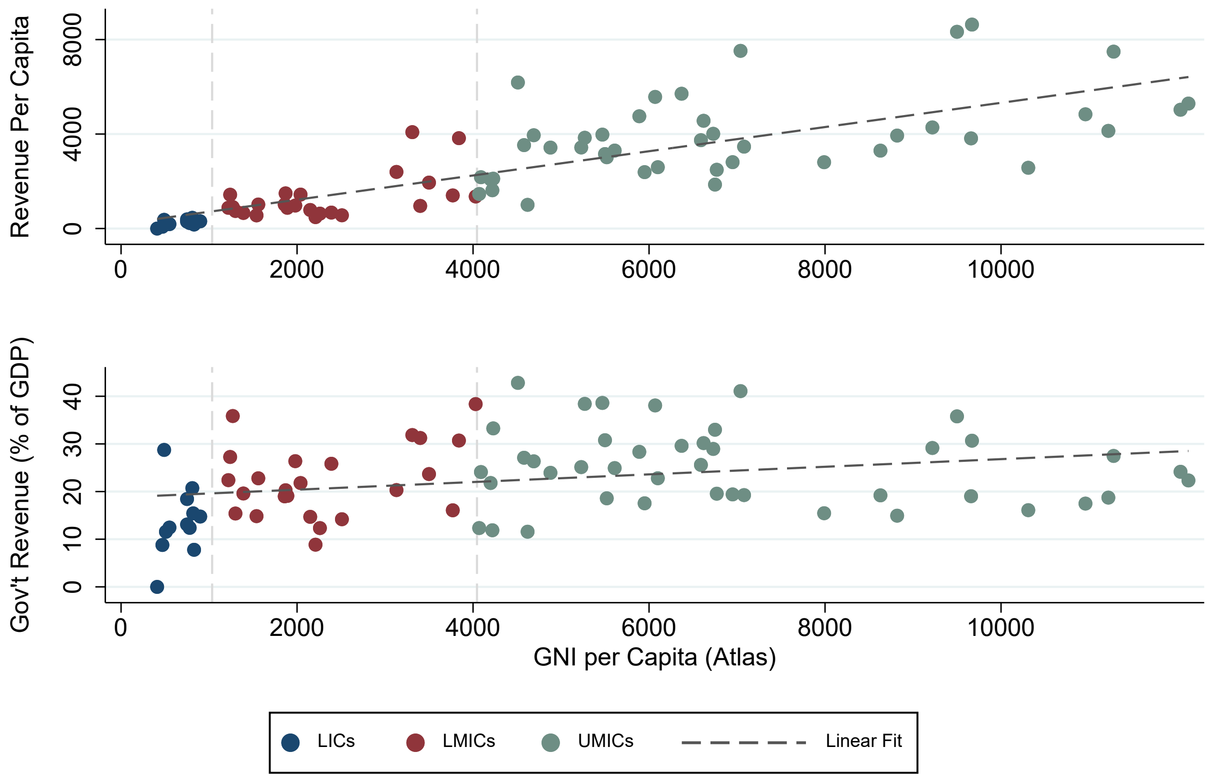
While FDI and remittances nonetheless entail a huge amount of resources, they are difficult to direct towards development investments even indirectly via regulation or tax incentives. Rather, private investments typically follow market demand and remittances often flow based on familial or inter-personal relationships – both which are less likely to serve people living in poverty and can potentially reinforce inequality, though Azizi (2021) finds that remittances are poverty and inequality reducing. As a result, the focus of my analysis is on domestic revenues – specifically the revenues generated by developing countries across income groups and their potential capability to redistribute resources to address its poverty gap. However, I also look at the relationship between other sources of financing and average income in more detail in Appendix 3.

Because domestic resources are the most critical in terms of volume and relevance to aid allocation, I first examine actual domestic resource mobilization in terms of government revenues per capita and as a percentage of GDP. Once the current picture of domestic resource generation is established, I compare the tax potential to the magnitude of the poverty challenge. I do this by conducting a test comparing the poverty gap across various poverty lines in a given country to the potential taxable domestic resources available to compute the marginal tax rate required for a country to finance its own poverty reduction and development.

Domestic resource mobilization

The poorest developing countries face two major problems in mobilizing domestic resources for development investments. First, their economic base is small by definition – a country at the LIC/LMIC threshold with an effective tax rate of 20 percent would only mobilize about $200 per capita, whereas a country at the LMIC/UMIC threshold would mobilize $800 per capita at the same tax rate or would only have to levy a tax of five percent to mobilize the same amount of resources. Second, countries at very low levels of development typically do not have high effective tax rates. This is more of a governance problem than a resource constraint, but it also limits the ability of low-income countries to mobilize resources. As a result, ODA is the largest source of financing for LICs (USAID 2015) and remains the most important resource for countries mobilizing less than $500 per capita (Development Initiatives 2013).

Figure 5. Government Revenues vs Average Income



Note: Government revenues plotted against GNI per capita.
Source: World Bank World Delevoment Indicators.

These two challenges are depicted in Figure 5, which shows both government revenue per capita[21] in the top panel and government revenue as a percentage of GDP in the bottom panel. The markers are plotted against GNI per capita on the horizontal axis with the dashed vertical lines showing the divide between the income groups. The markers in the top panel show government revenues per capita. There is very little variation in low-income countries, as they are all clustered at very low revenues per capita in the bottom left-hand corner with 20 of the 22 LICs with data mobilizing less than $500 per capita. The

[21] Some of the revenues per capita data were extrapolated forward from previous year's data.

median (mean) LIC mobilized just $237 ($288) per capita annually in government revenue. This compares to $1,031 ($1,578) for the median LMIC and $3,621 ($3,824) for the median UMIC. This means the median LMIC raises more than four times more government resources through taxation, and the median UMIC raises more than 15 times the median LIC. The dashed line shows the linear line of best fit, which shows a strong and positive correlation between average income and government revenue per capita, though the rate of increase slows at higher levels of average income for UMICs.

The differences are less dramatic when considering government resources relative to the size of the entire economy. Again, there is a positive relationship, which is shown by the dotted line representing the linear line of best fit in the bottom panel. This line is much less steep than revenues per capita, and there is more variation within the income groups. Tracing the markers to the vertical axis, half of the LICs (6 of 12) with data are clustered between 11 and 16 percent of GDP. The median (mean) LIC is 12.8 (13.7) percent, and the range is from zero percent (Somalia) up to 28.7 percent (Mozambique). LMICs have a higher median as well as a higher minimum value with all countries but one above 10 percent (Bangladesh). The median (mean) LMIC is 22.1 (27.9) percent of GDP, which is driven up by two outliers that are above 50 percent (in Timor-Leste at 67.6 percent and Kiribati at 125.6 percent) and are not shown in the figure. The inter-quartile range is from 16.1 percent to 30.7 percent. UMICs are higher still with a range of 11.6 percent at the minimum (Guatemala) up to 42.8 percent (Azerbaijan). The median (mean) UMIC is 24.5 (25.1) percent, and the inter-quartile range is 19.0 percent to 30.2 percent.

In general, LICs are not able to generate significant domestic revenues through taxation given their small economic base, but they also have lower effective tax rates. Revenue per capita grows with higher levels of average income, particularly for LMICs and relatively poor UMICs, but the rate of increase slows at higher levels of average income for UMICs. Effective tax rates also have a positive relationship with average income, but the increase is less pronounced as this does not depend on the size of the economy – only the capacity of the government to collect taxes and the willingness to impose taxes. From the donor perspective, the potential to redistribute resources is most important – even more than the actual domestic resources mobilized that I examine here – particularly when compared to the scale of the challenge. Now that I have established that actual resources mobilized are extremely low for countries at the lowest levels of development, I now turn to comparing the *potential* to mobilize resources as it compares to the financing gap needed – a more conceptual test of development need as opposed to the measures above that are influenced by the country's economic governance and capacity.

Potential redistributive capacity

The poverty gap is defined as the difference between an individual's level of consumption and the chosen poverty line. In other words, the amount of resources required to bring an individual up to the poverty line. This measure is typically presented as the "poverty gap index," which is the amount by which the average individual in a country falls below the poverty line expressed as a percentage of the poverty line averaged across all people in the country whether they are living below the poverty line or not. This is calculated using micro data from household surveys, so it is the difference between the level of consumption for an individual and the chosen poverty line summed for all people living in poverty. Then that number is averaged across the entire population (with people not living in poverty counting as zero). This provides a measure that makes it easy to calculate the shortfall in consumption from the poverty line for the average citizen. The formula for the poverty gap index (P_1) is presented below:

$$P_1 = \frac{1}{N} \sum_{i=1}^{H} \frac{Gi}{z}$$

Where N is the total population, H is the population living below the poverty line, G_i is the poverty gap (i.e., the difference between the chosen poverty line and individual *i*'s income or consumption), and z is the chosen poverty line. I am mainly concerned with the aggregate poverty gap for an entire country, so I convert the poverty gap index to an annual aggregate poverty gap.[22] This is roughly analogous to income per capita – or the annual poverty gap per capita in this case. Put differently, this is the amount of resources needed to close the poverty gap if the government could employ perfectly targeted and efficient transfers to people living in poverty to provide just the amount of resources needed to lift them to the chosen poverty line.

This sets me up to compare the total poverty gap to the amount of taxable domestic resources available in the economy, which can be considered in several different ways. While a closely related study of domestic financing gaps (Manuel *et al* 2018) considers each country's tax potential given various structural characteristics and limitations,[23] I follow Ravallion (2010) in considering a taxable resource base above a chosen level of consumption per day that excludes a "middle class". This effectively side-steps the morally objectionable implication of suggesting a marginal tax on the near poor to finance poverty reduction. I can then juxtapose the poverty gap against the resources of the "rich" above that line. Essentially, such a ratio of the poverty gap to the available resources implies a minimum MTR that would be required to eliminate the poverty gap with transfers. While the assumption of perfectly targeted and efficient direct transfers is unrealistic, it does provide us a sense of whether the elimination of poverty at a given poverty line is possible for a given country.

---

[22] While the poverty gap index is expressed as a proportion of the poverty line in consumption per day, I transform it into an annual poverty gap for the country expressed in PPP dollars. That is, I take the poverty gap index provided and multiply it by the associated poverty line (z in PPP dollars per day per person) to produce a poverty gap per person per day (Gi). Then I multiply that poverty gap by 365.24 to get an annual poverty gap per person. If I multiply that number by the total population, that gives the total poverty gap for the entire country.

[23] Manuel *et al* (2018) do not produce these estimates – rather, they split the difference between the IMF (Le *et al* 2012) and World Bank (Fenochietto and Pessino 2013) estimates of tax potential.

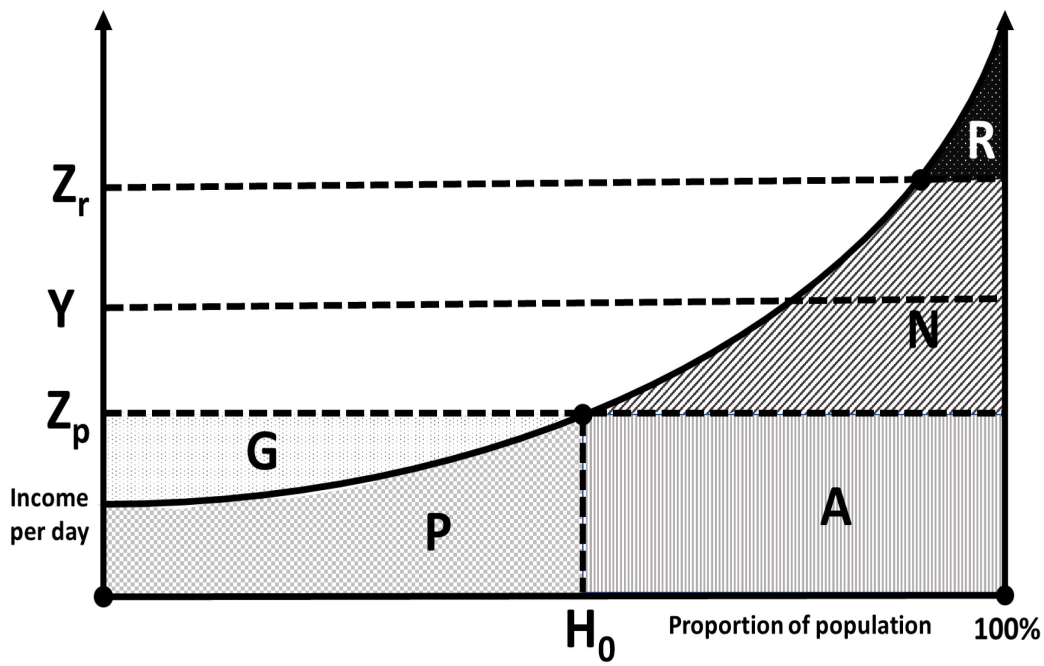
The diagram in Figure 6 depicts the set up conceptually. The vertical axis is consumption or income per person per day. The horizontal axis is a cumulative proportion of the population arrayed from poorest to richest (left to right). On the vertical axis, Z_p is the relevant poverty line – the extreme poverty line of \$2.15 per day for most of the analysis that follows below. *Y* is average consumption or income for the entire population, i.e., average income. Z_r is the consumption line that denotes the "rich". On the horizontal axis, H_0 is the proportion of the population consuming less than the poverty line, on average – this is the headcount poverty ratio. For sake of simplicity, I will refer to those consuming between Z_p and Z_r as the "middle class" and those consuming above *Zr* as the "rich".

Various areas on the diagram in Figure 6 are labeled to assist in explaining the calculation. *G* is the total poverty gap of a given country. *P* is the total consumption of the poor. Everything aside from *P* under the convex line (*A + N + R*) is the consumption of the non-poor. The area above the poverty line (Z_p) but below the curved line (*N + R*) is the total amount of resources above the poverty line, and *R* is the total amount of taxable resources of the "rich" above the chosen consumption line that separates the "middle class" from the "rich".

To operationalize the test, I compare the poverty gap (*G*) to the resources of the "rich" as defined by a consumption line of \$15 (*R*). A much lower bar would be the poverty gap as a proportion of the entire economy, i.e., the entire area below the convex line (*P + A + N + R*). This is the approach used by Sen (1981). Alternatively, I could look at the resources above the chosen poverty line (*N + R*) following Anand (1977). Instead, I choose to examine the resources above the higher consumption line (*R*). This follows Ravallion (2010) who uses a \$13 per day line based on the 2005 PPPs. Similarly, I choose a threshold of

\$15 per person per day for Z_r based on the same rationale and updated data.[24] The median poverty line for HICs is closer to \$25 per day, however, so rounding down to the lower threshold is a conservative choice as it maximizes the potential taxable resource base. While I follow Ravallion (2010) in the main text by only comparing the poverty gap to the resources above the "rich" consumption line, I also examine the other resource bases in Appendix 4, i.e., the total resources in the economy and the resources of those above the poverty line.

Figure 6. Diagram of Poverty Gap and Domestic Resources



All of the data used are from the World Bank's Poverty and Inequality Platform (PIP) database for 2019. PIP provides data for poverty rates and poverty gaps at the various poverty lines that I employ – \$2.15, \$3.65, \$6.85, and \$15 per day. It also provides a survey mean of consumption or income (depending on the survey) that I use as a proxy for total resources in the economy. I am able to transform these data to

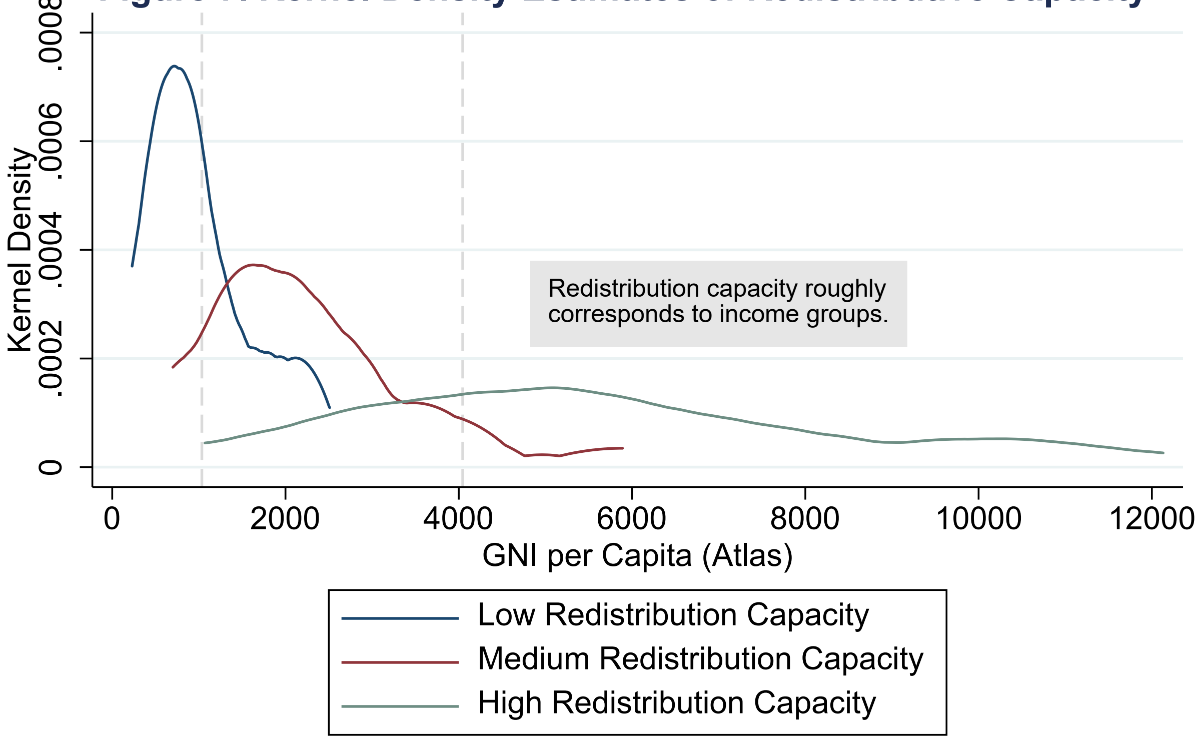
[24] Following Ravallion (2010), I use the US poverty line of \$25,750 for a family of four in 2019, which gives \$17.63 per day (i.e., (\$25,750/4 people) / 365.24 days = \$17.63 per person per day). I round down to use a conservative line and because the Poverty and Inequality Platform only provides estimates at \$5 intervals between \$10 and \$50.

calculate each country's poverty gap at various poverty lines and different resource bases, i.e., total resources, resources of those above the poverty line, and resources above the "rich" line.

When the poverty gap is presented as a proportion of the chosen resource base, this implies an MTR that would produce the resources necessary to fill the poverty gap through direct transfers that were perfectly targeted. This gives a sense of whether a country has sufficient domestic resources to finance its own development. If the MTR is above 100%, then it does not have enough resources available, even if they were all redistributed to the poor. Short of 100%, I must choose an arbitrary MTR that would be considered prohibitively high, so I follow Ravallion (2010) in choosing 60 percent. I define a "low" capacity to redistribute as above 60 percent of the resource base, "medium" capacity as between 10 and 60 percent, and "high" redistributive capacity as below 10 percent of the resources available.

Figure 7 shows the distribution of redistributive capacity by GNI per capita for a $2.15 per day poverty gap and a resource base of only the rich (i.e., above the $15 per day line). Starting with the low-capacity countries (above an MTR of 60 percent), every single one of these countries is a LIC (22 countries) or relatively poor LMIC (9). The richest country with a low capacity to redistribute is Papua New Guinea, an LMIC at $2,510 GNI per capita in 2019. The medium capacity countries are primarily LMICs (15 of the 19 countries), though there are also three LICs (The Gambia, Burkina Faso, and Mali). The richest country (and only UMIC) with a medium capacity to redistribute is Belize at $5,890 GNI per capita. Finally, the countries with a high redistributive capacity are mostly UMICs (52 of 74 countries), though there are also some LMICs (22) but no LICs. The poorest country with a high redistributive capacity is Tajikistan with $1,070 GNI per capita. To summarize, the capacity to redistribute resources to close the poverty gap is strikingly well aligned with the income group classifications.

Figure 7. Kernel Density Estimates of Redistributive Capacity



Note: Kernel density estimates of redistribution capacity distributed across GNI per capita (Atlas).
Source: World Bank Poverty and Inequality Platform, World Develoment Indicators.

The estimates I have presented thus far have used the same set of assumptions: first, that the extreme poverty line of $2.15 per day was the most relevant threshold to measure the poverty gap; and second, that the potential taxable resource base was only above the "rich" threshold of $15 per day. It is possible to use different assumptions for both of these parameters, however. Regarding the choice of poverty line, the World Bank now determines three absolute poverty lines, which are updated when a new set of PPPs are released. The most recent poverty lines were set in October 2022 using the new 2017 PPPs, and the 2019 poverty estimates used here were released at the same time. The $2.15 per day poverty line is the median national poverty line of LICs, whereas the $3.65 poverty line is the median national poverty line of LMICs and the $6.85 per day poverty line is the median for UMICs. Therefore, for my purposes of analyzing the redistributive capacity across income groups, it is arguably more appropriate to use the absolute poverty line specific to the relevant income group being analyzed.

Second, regarding the choice of the resource base, Ravallion (2010) innovated on previous attempts by exempting a "middle class" from the taxable resource base. However, it is possible to instead compare the poverty gap to the resources of the non-poor (Anand 1977) or the entire economy to determine the "country's potential ability to meet the challenge of poverty" (Sen 1981, p. 190). While I am sympathetic to Sen's approach given that it echoes my earlier arguments for average income as the most useful indicator of country need,[25] on the other hand, it is not as fit-for-purpose as the more conservative approaches from the perspective of examining potential redistributive capacity. To elaborate, the implied approach to taxation produced by Sen includes a tax on the poor, which would then be returned to them. Therefore, I find it more defensible to consider only the taxable resources of the non-poor (*N + R* in Figure 6 above) as a complement to the resources of only of rich (*R*). Especially at lower levels of development where there are not many people living above the "rich" threshold, it is realistic to consider all of the non-poor as part of the tax base to calculate the necessary marginal tax rate on income above the poverty line. I include only the non-poor resource base in my analysis here, though I also include the Sen (1981) approach of including the entire economy's resources in Appendix 4.

Table 1 shows the implied marginal tax rates on two different resource bases to close the poverty gap at the relevant poverty line for that income group. I have already shown above that most LICs have a very low capacity to redistribute resources to close the poverty, and these summary statistics reinforce that.

[25] In addition, this approach has the political advantage of sharing the tax burden. This helps to alleviate the political economy problem of only levying a tax on the rich.

When only taxing the "rich", the implied tax rate outstrips the available resources for two thirds of LICs (16 of 24 countries), i.e., the poverty gap is greater than the total resources above the "rich" threshold, and seven out of every eight LICs (21 of 24) have a low capacity to redistribute. The median LIC's implied marginal tax rate is approximately twice the available resources of the rich above the $15 per day line. However, when considering the potential to tax all the non-poor's resources above the poverty line, the capacity to redistribute improves dramatically for many countries – this is at least partially because LICs have so few people living above the $15 per day line. Fewer than one third of LICs (7 of 24 countries) fall into the "low" capacity to redistribute category with an MTR above 60 percent when the tax base is expanded to include all resources above the poverty line. This is an extraordinarily low bar, however, and it is difficult to ask LICs to tax those just above the $2.15 poverty line to finance transfers to those just below that poverty line. This was the case that Ravallion (2010) made forcefully and is the reason why he chose a significantly higher threshold in "exempting" the middle class.

Table 1. Marginal Tax Rates Required to Close Poverty Gap, 2019

Income Group	Taxable Resource Base	$2.15/Day Poverty Line			
		Mean	**25th Percentile**	**Median**	**75th Percentile**
Low	**Non-Poor**	47.7%	76.5%	17.3%	5.5%
	Only Rich	1304.6%	1575.1%	201.1%	78.6%
		$3.65/Day Poverty Line			
Lower Middle	**Non-Poor**	28.9%	36.9%	16.6%	2.5%
	Only Rich	230.7%	292.8%	94.4%	17.2%
		$6.85/Day Poverty Line			
Upper Middle	**Non-Poor**	13.2%	12.8%	6.5%	1.8%
	Only Rich	34.0%	31.8%	14.1%	3.8%

Note: Total poverty gap at the stated poverty line as a proportion of the resource base (as measured by survey consumption or income per capita in PPP terms). Percentile ranks are determined by the MTR within each income group (and not the resource base).
Source: World Bank Poverty and Inequality Platform.

LMICs have a greater capacity to redistribute than LICs, and this capacity is much greater if taxing all of the non-poor according to the $3.65 per day poverty line. If taxing only the rich, most LMICs have a low capacity to redistribute (29 of 47), and the median is nearly 100 percent of the rich's resources. Only three LMICs have a high capacity to close the $3.65 per day poverty gap when taxing only the rich (Ukraine, Tunisia, and Bolivia). Like LICs, the analysis is much different if I consider taxing all of the resources above the $3.65 per day poverty line. The median LMIC would have to tax all non-poor at an MTR of 16.6 percent, and the average LMIC would have to tax about 29 percent. While there are only eight LMICs with a high capacity to redistribute when taxing all non-poor resources, about seven of ten LMICs (33 of 47) have a medium capacity to redistribute, which means they have an MTR above ten percent but less than 60 percent. While less problematic than taxing those living on $2.16 a day, the $3.65 per day poverty line is still fairly low (at around $1,300 annually), so it is still questionable whether those living on less than $4 per day should be taxed to finance poverty reduction efforts.

Finally, UMICs have a high capacity to redistribute regardless of the resource base. Even if only taxing the rich, more than two in five (22 of 53 UMICs) have a high capacity to redistribute. The median UMIC would have to mobilize about a 14 percent MTR to generate enough revenues to transfer away the poverty gap at the $6.85 poverty line. The mean MTR on resources above the "rich" threshold is 34 percent. However, if taxing all of the resources in the economy above the $6.85 poverty line, the capacity to redistribute is much higher and all but two countries have either a high or medium capacity (Belize and Indonesia, both below $6,000 GNI per capita), i.e., less than an implied 60 percent MTR. Two thirds of UMICs (35 of 53 countries) have a high capacity to redistribute when taxing all resources above the $6.85 poverty line, and the median (mean) UMIC would have to tax 6.5 (13.2) percent of non-poor resources. In contrast to the lower poverty lines, $6.85 per day is nearly half of the $15 per day US poverty line used here, and it is more defensible to tax people living on more than that amount.

As in previous sections, I conclude by examining the level of average income above which most (or all) countries achieve a high capacity to redistribute. Returning to the redistribution capacity to end extreme poverty with resources above the “rich” threshold, there are no UMICs with a low capacity to redistribute and only one UMIC with a medium capacity to redistribute in Belize (with a 10.2 percent MTR). Furthermore, only four other UMICs would have to employ an MTR above three percent – Georgia, Belize, Botswana, and South Africa – and it would be between 3.1 and 3.4 percent. Notably, these four countries with the lowest capacity to redistribute are all below the IBRD line. This pretty clearly suggests that any country above the average income of Belize, around $6,000, should have the ability to self-finance poverty eradication, which is slightly lower than the IBRD line suggested by the analysis related to the magnitude of the challenge (i.e., consumption and multi-dimensional poverty). Notably, Ravallion (2010) conducts a similar exercise and cites a line of $4,000 in consumption per capita (using 2005 PPP data) at which the MTR is below one percent, on average. While this is higher than my $6,000 line in current 2019 dollars, he also uses a lower bar – all countries above that $4,000 threshold have an MTR of less than one. Emulating the same test, I find that the country with an MTR above one with the highest average income in Atlas terms is South Africa at $6,730. This again suggests that all countries above the IBRD line should be able to easily finance poverty reduction.

In sum, the necessary rates of taxation to close the $2.15 per day poverty gap are excessively high for LICs and most LMICs but easily manageable for most UMICs. This complements the finding that there are significantly fewer government resources mobilized by LICs and LMICs for development investments and established that it is not solely due to lower effort or lower capacity in terms of tax administration – it is mainly a problem due to a lack of resources. The capacity to redistribute resources to close the extreme poverty gap is strikingly well aligned to the income group classifications. When considering the potential to tax all the non-poor’s resources above the poverty line, the capacity for LICs to redistribute improves dramatically for many countries. LMICs also have a greater capacity to redistribute if taxing all of the non-poor even at the higher $3.65 per day poverty line. The policy implication is that especially

LICs and most LMICs could benefit greatly from grant assistance, whereas most UMICs can finance their own poverty reduction and development. UMICs have a high capacity to redistribute regardless of the resource base, and all countries above the IBRD line are able to self-finance poverty eradication with an MTR on their rich population of less than one percent.

Policy implications and conclusion

I conclude by synthesizing my findings, discussing the policy implications of my findings, and providing suggestions for how donors might be more responsive to this evidence in employing a more needs-based allocation model.

I have shown that both extreme poverty and multi-dimensional poverty are strongly associated with average income, and there are significant differences in country-level outcomes across income groups. This makes a strong case that LICs in particular – but also LMICs – are in much greater need of external financing for development investments and that aid should be cut off above the IBRD line. Given that the countries with the lowest average incomes have both the highest rates of poverty and the lowest rates of long-run growth (by definition), they are obvious candidates for assistance intended to promote poverty reduction and economic growth. This is further supported by the observation by Kenny (2020) that the marginal utility of income is greatest for those living in countries where poverty is widespread. Furthermore, if a policymaker also wants to explicitly focus on multi-dimensional poverty beyond just consumption poverty, the evidence leads to the same implication. That is, multi-dimensional poverty is also greatest in the countries with the lowest average income, and Kenny's (2020) argument still holds – the poorest households are most likely to expend marginal income to fulfill basic human needs and acquire absolute necessities.

However, some policymakers might argue that there is heterogeneity across multi-dimensional poverty and that aid should focus on one (or more) specific social outcomes that they deem especially critical for long-term development, such as primary education. While this thinking would still lead back to many of the same LICs that perform worst on measures of multi-dimensional poverty, there is also a strong case for investing in broader economic growth. Pritchett and Lewis (2022) find that economic growth is both necessary and sufficient to achieve basic human needs, and my analysis supports this. Therefore, a narrow focus on aid provision to a particular sector without considering the broader economic growth narrative may not be the highest return investment. Rather, donors should put their resources into the highest return sectors and investments in terms of the potential economic returns with the recognition that broader economic growth and rising average incomes will also benefit other sectors than the ones on which they are focused. This is akin to the "growth diagnostics" approach espoused by Hausman, Rodrik, and Velasco (2008) that has been operationalized by various donors, such as the World Bank, USAID, DFID, and especially MCC. This type of "constraints to growth" analysis lends a disciplined and data-driven approach to sector prioritization that leads to a focus on broader economic growth objectives with the highest potential returns, rather than narrow sector-based objectives.

Turning to domestic resource mobilization, I found that government revenues per capita are extremely low in LICs, but they get progressively larger as an economy grows and reaches higher levels of average income. I also find that the necessary rates of taxation to close the poverty gap with direct transfers are excessively high for LICs and most LMICs but easily manageable for most UMICs. This is both because the poverty gap is much larger in LICs and LMICs and because the domestic resources base is much larger in UMICs. This all suggests that countries are increasingly able to finance their own development as they reach higher levels of average income, particularly as they reach UMIC status and certainly once they cross the IBRD line. These findings are consistent with the strategic vision of many donors that look to transition countries away from grant assistance towards financing their own development. For instance, USAID has focused in recent years on both “localization” and “self-reliance”, which have been strategic pushes towards putting countries in charge of their own development and transitioning them away from a dependence on foreign aid, respectively. My findings suggest that most grants should be targeted to LICs and LMICs as UMICs are largely “self-reliant”. Dassinayake *et al* (2020) make a related point arguing that ODA should flow to LICs and LMICs, and grants going to UMICs should be approached differently and have a large, expected impact in address an outstanding development challenge, such as working with under-developed regions or marginalized groups.

This chapter brought together the resource availability issue with observations related to the magnitude of the development challenge, and putting it all together, the two different components of country need both point in the same direction – with the same policy implication of a strong focus on allocating assistance towards the poorest countries in terms of average income, i.e., LMICs and particularly LICs. Put simply, most grant financing should go to the very poorest countries and should be cut off around the IBRD line in terms of average income. While some relatively poor UMICs may have some persistent poverty that might warrant donor attention, the IBRD line offers a data-driven potential option as a threshold for cutting off assistance. I find that a reasonable cut-off is around $6,000 to $7,000 GNI per capita – roughly the IBRD line – when considering whether countries can be expected to eliminate poverty and/or self-finance their own poverty eradication. At the very least, grants going to UMICs should be approached differently and have large impacts in addressing outstanding development challenges, while starting to transition the diplomatic engagement away from an aid relationship.

This type of analysis is rarely applied to policy decisions regarding how to allocate foreign assistance across income groups, yet donors use (and sometimes ignore) these income thresholds in operational ways with huge implications. For instance, the World Bank employs an IDA threshold at the lower end of the LMIC income group to trigger graduation away from grant assistance (but the graduation process sometimes plays out over a decade or more). MCC is statutorily required to apply a filter for country need by limiting its pool of potential partners to just LICs and LMICs (but is considering expanding its candidate pool to includes UMICs). This analysis provides supporting evidence to continue (and double down on) disciplined, needs-based approaches and provides strong evidence that they should be employed more widely and strictly by donors. The one-two punch of the magnitude of the challenge complemented by limited domestic resources makes a potent argument that LICs and LMICs should be strongly favored when it comes to the allocation of grant assistance with a potential cut-off of assistance around the IBRD line, if not below.

Appendix 1. Total number of people living in poverty

This appendix provides a side-by-side comparison of the countries with the highest poverty rates and total number of people living in poverty – there is some overlap, but both lists are mostly populated by LICs, especially at the top. Table A1a shows the countries with the highest poverty headcount ratio using the international poverty line of $2.15 per day. This is simply the proportion of the population consuming less than $2.15 per day, on average. The table shows that nine of the ten countries with the highest poverty rates are LICs, and the countries with the highest prevalence of extreme poverty are Madagascar, Burundi, and South Sudan. An LMIC doesn't appear on the list until you get to the tenth country, Zambia, which is a highly unequal and relatively poor LMIC. There are more LMICs in the rest of the table, and there are seven LMICs of the twenty countries, but not UMICs. This is presented for the purposes of comparison to the total number of people living in extreme poverty.

Table A1a. Countries with Highest Headcount Ratios ($2.15/day poverty), 2019

	Country	Percentage	Income Group
1	Madagascar	79.0%	LIC
2	Burundi	73.6%	LIC
3	South Sudan	71.9%	LIC
4	Malawi	69.1%	LIC
5	CAR	67.6%	LIC
6	Somalia	65.3%	LIC
7	Syria	64.2%	LIC
8	Mozambique	63.2%	LIC
9	DRC	62.3%	LIC
10	Zambia	61.1%	LMIC
11	Yemen	58.7%	LIC
12	Rep. of Congo	50.8%	LMIC
13	Niger	50.1%	LIC
14	Rwanda	44.2%	LIC
15	Tanzania	43.1%	LMIC
16	Uganda	40.5%	LIC
17	Zimbabwe	39.8%	LMIC
18	Eswatini	34.2%	LMIC
19	Lesotho	33.7%	LMIC
20	Angola	32.2%	LMIC

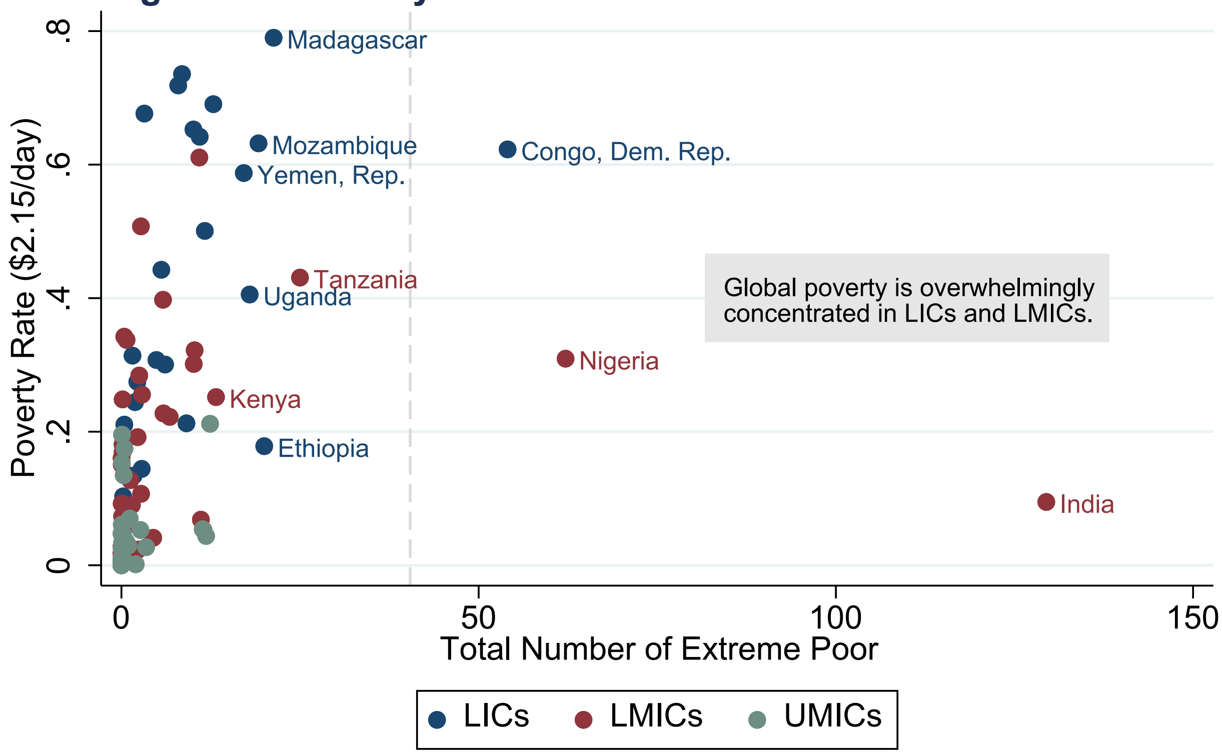
Table A1b. Countries with Largest Populations Living in Poverty($2.15/day poverty), 2019

	Country	Millions	Income Group
1	India	129.0	LMIC
2	Nigeria	62.2	LMIC
3	DRC	54.1	LIC
4	Tanzania	25.0	LMIC
5	Madagascar	21.3	LIC
6	Ethiopia	20.0	LIC
7	Mozambique	19.2	LIC
8	Uganda	18.0	LIC
9	Yemen	17.1	LIC
10	Kenya	13.2	LMIC
11	Malawi	12.9	LIC
12	South Africa	12.4	UMIC
13	Indonesia	11.9	UMIC
14	Niger	11.7	LIC
15	Pakistan	11.5	LMIC
16	Brazil	11.4	UMIC
17	Bangladesh	11.1	LMIC
18	Syria	11.0	LIC
19	Zambia	10.9	LMIC
20	Angola	10.2	LMIC

As mentioned in the main text, the total number of people living in $2.15 per day poverty is less well correlated with average income as it depends heavily on population. Table A1b shows the twenty countries with the largest numbers of people living on less than $2.15 per day. India has the largest poor population by far and Nigeria and DRC are second and third but have less than half that amount each. However, India is rapidly reducing their poverty rate, so this number will continue to decrease dramatically, whereas Nigeria and DRC's poverty rates are more persistent. There are six LICs, four LMICs, and no UMICs in the first ten countries, though there are three UMICs further down the list in South Africa, Indonesia, and Brazil. In sum, the first ten countries account for about 380 million people and all twenty countries add up to almost half a billion people (494 million). Given that the total number of extreme poor was estimated at 648 million in 2019 (World Bank 2022), that means that the first ten countries accounted for nearly six in ten extreme poor globally (58.6 percent), and the full list accounted for more than three quarters of global poverty (76.2 percent). In other words, global poverty is highly concentrated in this short list of countries that is primarily LICs and LMICs.

While there is significant overlap between the two lists in Table A1, there is not perfect overlap because some of the poorest countries have relatively small populations (e.g., Burundi) and some of the UMICs with a large number of poor do not have relatively high poverty rates (e.g., Brazil). This relationship is show in Figure A1, which displays a scatter plot of the poverty rate on the vertical axis versus the total number of extreme poor on the horizontal axis. The countries with the ten highest numbers of extreme poor are labeled, and they are all LMICs (red markers) or LICs (blue markers), as shown above. There is no clear relationship, except that there are very few UMICs with large nubmers of extreme poor. Setting aside the case of India (which is historically similar to China in that it had large numbers of poor but reduced that number rapidly over time), there is a trend that all of the countries with the largest number of poor also have a high poverty rate. Besides India, only Ethiopia has a poverty rate below 20 percent of the ten coutnries labeled with the greatest numbers of extreme poor. The implication here is that targeting the largest number of poor is a crude instrument – while it obviously helps to identify countries accouting for large proportions of global poverty, it does not necessarily identify the countries with the highest poverty rates. This is an important distinction as there is a tendency for policymakers to prioritize highly visible countries with the largest numbers of poor (like India) instead of more precisely targeting the poorest countries with the greatest need (like Burundi).

Figure A1. Poverty Rate vs Total Number of Poor



Note: Poverty rate at $2.15 per day plotted against total number of extreme poor in 2019.
Source: World Bank Poverty and Inequality Platform, World Develoment Indicators.

Appendix 2. Correlations between development outcomes of interest

In general, most measures of development progress are well correlated with each other. In particular, most measures of development progress are well correlated with average income. Table A2 shows the partial correlations between the various development outcome indicators used in the main text. Throughout this chapter, I use both GNI per capita in Atlas terms as a measure of broad development progress alongside household consumption measured in PPP terms. Table A2 shows that GNI per capita is almost perfectly correlated with mean household consumption at the country level, so this should not be problematic for the analysis conducted above. The extreme poverty rate has the lowest coefficient at -0.37. The rest of the indicators have a coefficient of .40 to .66 with respect to their relationship with GNI per capita, which suggests that these outcomes all move together, i.e., broad development progress on average income and multi-dimensional outcomes are mutually reinforcing.

Table A2. Pairwise correlations of various development indicators

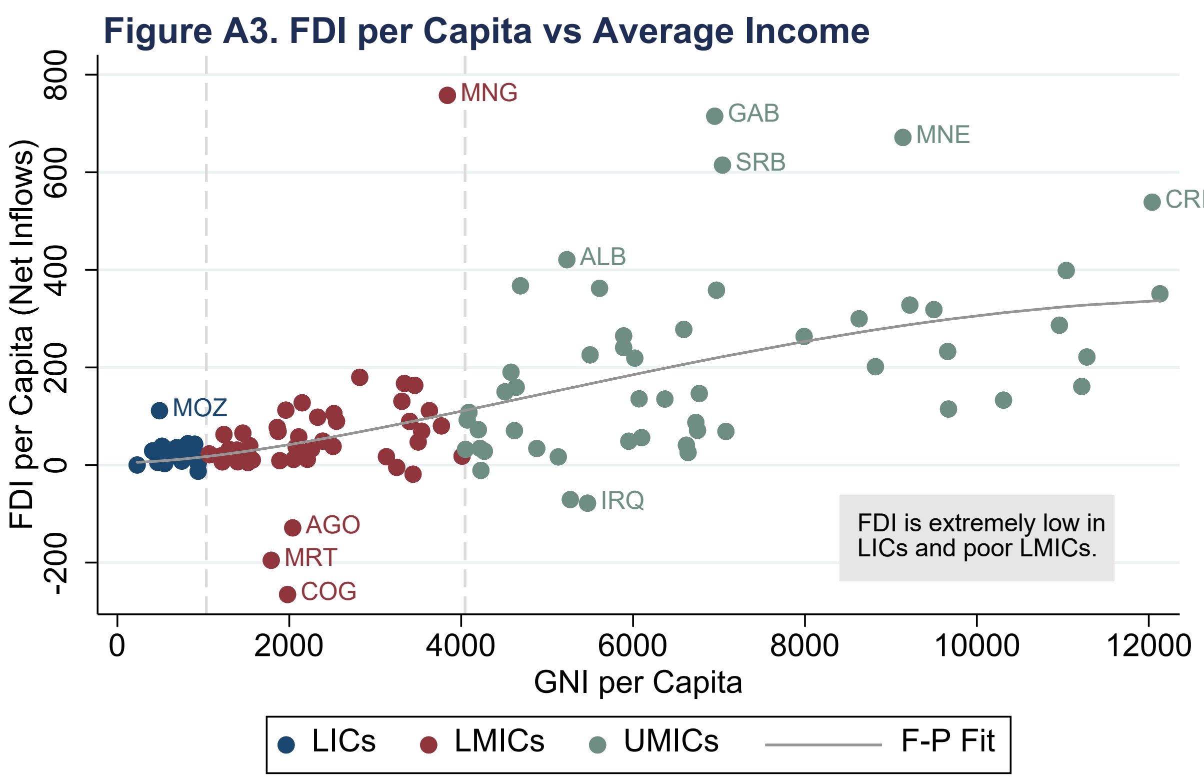
Variables	GNI per Capita	Mean Consumption	$2.15/Day Poverty	Access to Energy	Access to Water	Primary Education	Life Expectancy	Infant Mortality	Access to Energy
GNI per Capita	1.00								
Mean Consumption	0.98	1.00							
Extreme Poverty Rate	-0.37	-0.40	1.00						
Access to Energy	0.40	0.43	-0.85	1.00					
Access to Water	0.46	0.49	-0.82	0.90	1.00				
Primary Education	-0.58	-0.64	0.56	-0.77	-0.74	1.00			
Life Expectancy	0.66	0.68	-0.71	0.80	0.82	-0.64	1.00		
Infant Mortality	-0.51	-0.52	0.71	-0.84	-0.85	0.74	-0.92	1.00	
MPI Rate	-0.64	-0.69	0.72	-0.90	-0.89	0.93	-0.77	0.85	1.00

Source: World Bank Poverty and Inequality Platform and World Development Indicators.
Note: All data from 2019 and only for developing countries. GNI per capita is in Atlas terms; mean consumption is in PPP terms; extreme poverty rate is $2.15/day poverty; multi-dimensional outcomes are all the same as in the main text.

Furthermore, it is worth noting that the correlation between MPI and the extreme poverty rate is extremely high at .72. Again, this shows that my primary measure of consumption in monetary terms is closely associated with other multi-dimensional poverty outcomes. Not surprisingly, the MPI rate is strongly correlated with the multi-dimensional poverty outcomes, as several of the individual indicators are included as one of the MPI's components.

Appendix 3. Other Sources of Development Finance

The main text focuses primarily on government revenues because they are arguably the most important source of development finance. This is because they are much larger than international resource flows to developing countries and because tax policy is within the policy domain of the country partner governments. Furthermore, while private capital and remittances entail a huge amount of resources, they are difficult to direct towards development investments. Nonetheless, while private investments typically follow market demand and remittances flow based on familial or inter-personal relationship, both are critical potential sources of development finance. As a result, I also look at the relationship between other sources of financing and average income in more detail in this appendix.

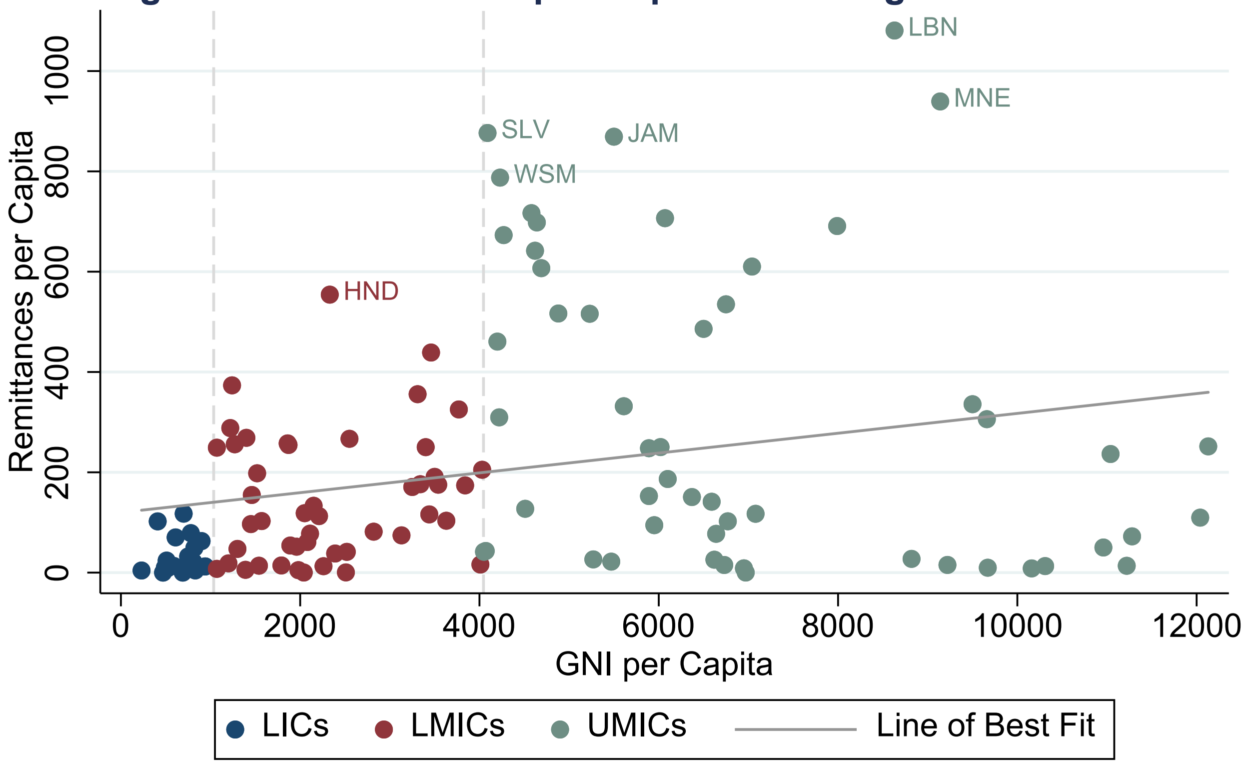


Note: Foreign direct investment (net inflows) plotted against GNI per capita (Atlas) in 2019.
Source: World Bank World Develoment Indicators.

Table A3 shows the distribution of FDI per capita versus average income measured in GNI per capita in Atlas terms. The measure of FDI is net inflows (and not outflows), which is the value of direct private investment made by non-residents of the economy. The net nature of the measure captures both investment and dis-investment, so a negative value is possible and implies that dis-investment was greater than investment by non-residents during that period. The table shows that FDI per capita is extremely low for LICs and most poor LMICs with a few LMICs that are significantly negative (Angola, Mauritania, and Republic of the Congo). The median LIC receives just $21 per capita, and the median LMIC just $38 per capita. Mozambique is the only LIC with an FDI per capita greater than $50 at $111, and only one LMIC was above $200 per capita, Mongolia at $758.

There is much more variation in UMICs, including two outliers that are well above $1,000 per capita, The Maldives at $1,810 and Guyana at $2,166. The median UMIC receives $161 in FDI per capita, which is more than four times the median LMIC and nearly eight times the median LIC. While there are three UMICs with net dis-investment (Iraq, Namibia, and Samoa), nearly two thirds of UMICs with data (35 of 53) receive FDI per capita of greater than $200 in 2019. However, there are a number of relatively poor UMICs that are just above the LMIC/UMIC threshold that receive relatively little FDI per capita that is closer to the median LMIC than the median UMIC. More specifically the eleven poorest UMICs with data on FDI are all below $200 per capita, and all but one is below the UMIC median of $161 per capita. Four of the poorest seven UMICs are below the LMIC median of $38 per capita – Algeria ($233), Sri Lanka ($34), Samoa (-$11) and West Bank and Gaza ($28). This provides some support to the view that there are a set of relatively poor UMICs that have difficult mobilizing private capital and access international capital markets and therefore should be extended concessional lending, if not grants.

Figure A4. Remittances per Capita vs Average Income



Note: Remittances per capita plotted against GNI per capita (Atlas) in 2019.
Source: World Bank World Develoment Indicators.

Table A4 shows remittances received per capita versus average income for 2019.[26] Again, remittances per capita are extremely low for LICs compared to the rest of developing countries. The median LIC received remittances of $15 per capita, while the median LMIC received $116 – nearly eight times greater. Only one LIC, The Gambia, received remittances greater than the median LMIC, and all but two LICs (18 of 20) with data on remittances received less than $100 per capita. Though several UMICs had virtually no remittances (e.g., Angola and Papua New Guinea receive less than $1 per capita), there is much more variation within this income group with nearly three in ten LMICs with data (14 of 47) receiving more than $200 per capita, on average. Honduras received the most remittances among LMICs with $554 per capita. UMICs were even more heterogeneous – again, four countries received less than

[26] Remittances are defined as personal transfers – either in cash or in kind – and compensation from nonresidents. Compensation includes seasonal and short-term workers employed in an economy where they are not resident.

$10 per capita. However, the median UMIC was more than double the median LMIC at $236 per capita. Two countries received more than $1,000 per capita in Lebanon ($1,080) and Tonga ($1,821), which is an outlier not shown in the figure.

Where there is a trend, it appears that relatively poor UMICs receive more remittances than relatively rich UMICs. That is, most UMICs near the UMIC/HIC threshold receive relatively few remittances, whereas many of the highest remittance receiving UMICs were very close to the LMIC/UMIC threshold. This pattern does not necessarily hold for LMICs and LICs, however. The trend across the lower income groups is that remittances are more closely correlated with average income. LICs receive very little remittances overall, while LMICs receive more but not as much as UMICs. Beyond the general observation that LICs have very few resources across all sources of development finance, there are no clear implications for aid allocation given the inter-personal nature of remittances.

Appendix 4. Marginal Tax Rates Under Different Assumptions

In the main text, I only show a table of MTRs for taxing the non-poor and only the rich at the poverty line relevant to the income group (i.e., $2.15 per day for LICs, $3.65 per day for LMICs, and $6.85 per day for UMICs). In Table A3, I show the full set of MTRs for the mean, median, and inter-quartile range for each income group for whether they tax the total economy (including the poor), only the non-poor (above the relevant poverty line), or only the rich (above $15 per day) to close the poverty gap for all three poverty lines. This is a richer set of information for the reader that might be interested in checking the MTR for different resources bases or for the poverty gap at different poverty lines.

Table A3. Marginal Tax Rates Required to Close Poverty Gap, 2019

Income Group	Taxable Resource Base	$2.15/Day Poverty Line			
		Mean	25th Percentile	Median	75th Percentile
Low	Total Economy	14.83%	25.95%	8.24%	2.81%
	Non-Poor	47.68%	76.46%	17.25%	5.50%
	Only Rich	1304.62%	1575.09%	201.09%	78.57%
Lower Middle	Total Economy	2.47%	3.51%	0.72%	0.12%
	Non-Poor	4.55%	6.48%	1.15%	0.15%
	Only Rich	48.49%	44.39%	12.68%	0.87%
Upper Middle	Total Economy	0.20%	0.15%	0.03%	0.00%
	Non-Poor	0.25%	0.18%	0.03%	0.00%
	Only Rich	0.74%	0.50%	0.11%	0.00%

Income Group	Taxable Resource Base	$3.65/Day Poverty Line			
		Mean	25th Percentile	Median	75th Percentile
Low	Total Economy	47.28%	73.18%	33.18%	17.38%
	Non-Poor	278.48%	461.30%	108.40%	49.85%
	Only Rich	3880.86%	4977.04%	808.36%	453.77%
Lower Middle	Total Economy	10.50%	16.50%	6.95%	1.30%
	Non-Poor	28.93%	36.89%	16.61%	2.53%
	Only Rich	230.71%	292.81%	94.39%	17.21%
Upper Middle	Total Economy	0.94%	0.78%	0.39%	0.05%
	Non-Poor	1.40%	1.05%	0.54%	0.06%
	Only Rich	4.49%	3.52%	1.60%	0.19%

Income Group	Taxable Resource Base	$6.85/Day Poverty Line			
		Mean	25th Percentile	Median	75th Percentile
Low	Total Economy	139.11%	200.53%	108.30%	70.50%
	Non-Poor	2166.55%	3421.18%	772.17%	638.66%
	Only Rich	10806.68%	14429.13%	3274.98%	1589.43%
Lower Middle	Total Economy	43.36%	61.12%	40.03%	16.85%
	Non-Poor	253.46%	365.89%	159.69%	57.55%
	Only Rich	987.12%	1328.32%	508.05%	227.96%
Upper Middle	Total Economy	5.78%	6.18%	3.58%	1.06%
	Non-Poor	13.16%	12.83%	6.46%	1.77%
	Only Rich	34.03%	31.85%	14.06%	3.85%

Note: Total poverty gap at the stated poverty line as a proportion of the resource base (as measured by survey consumption or income per capita in PPP terms). Percentile ranks are determined by the MTR within each income group (and not the resource base).

Source: World Bank Poverty and Inequality Platform.

Chapter 3. Macro Evidence on Effectiveness

This chapter addresses the following question: Is foreign aid more effective in promoting economic growth in better-governed and more democratic countries? The concept of performance-based allocation means that foreign assistance is allocated to countries that are better governed, as it is believed that they are more likely to use it well. However, this approach inherently allocates aid away from countries that are poorly governed yet still have great need (Bourguignon and Platteau 2015). By choosing to work with well-governed countries, it is believed that aid will be more effective and achieve greater development impact (Burnside and Dollar 2002). That is, for any given investment, allocating aid resources to better-governed countries will result in a greater return on investment for each dollar of assistance in terms of the amount of economic activity generated – even if the marginal utility of each dollar might be higher in a poorer but less-well-governed country. This is known as the “aid-policy-growth” question or the “conditional” strand of the aid-growth literature, and the trade-offs associated with a performance-based approach motivate my question of whether a more selective approach to aid allocation results in improved development outcomes at the macro level.

The related economic literature generally falls into three strands: (a) the “conditional” strand that claims that aid only impacts growth in well-governed countries; (b) the “unconditional” strand that finds that aid impacts growth in all cases; and (c) the “null” strand that there is no relationship between aid and growth. I confirm the unconditional strand in my analysis, but I am most interested in the conditional strand, as it has greatly influenced donor policy regarding aid allocation. To answer the conditional question, I use Galiani *et al* (2017) as my starting point by exploiting a similar identification strategy whereby after a country crosses the arbitrary International Development Association (IDA) threshold it causes donors to decrease their aid allocations to that country. This creates a natural experiment that allows me to isolate the effects of exogenous variation in aid flows on changes in economic growth. To test the conditional aid-policy-growth question, I examine whether there is a change in the aid-growth relationship for better-governed or more democratic countries. To do this, I construct a set of country-specific indicators reflecting whether a country passed or failed various elements of MCC's country scorecard at the time of crossing the IDA threshold, e.g., control of corruption. As a result, I am able to test the aid-growth relationships of passers versus failers, i.e., relatively well governed countries versus relatively poorly governed countries.

The baselines OLS estimates for the fixed effects model are in line with the existing literature, but an unexpected pattern starts to emerge when I include interaction terms that effectively split my sample between scorecard passers and failers. If there is any pattern at all, it is that worse-governed and less democratic countries achieve greater growth outcomes from foreign aid. That is, better-governed and more democratic countries tend to have a weaker relationship in terms of aid's effect on growth. These baseline OLS estimates with scorecard interaction terms suggest it is possible that the relationship between aid and growth may be stronger for worse-governed and less democratic countries. This establishes the baseline for comparison to the IV estimates, but it is well recognized that the allocation of assistance may be endogenous to other factors, such as conflict or natural disaster. Since aid is often allocated directly in response to negative (growth) shocks or long-term economic stagnation, naïve OLS estimates may be biased downwards.

I then move on to the IV models. In testing the first stage for the full sample, I find that there is a statistically significant decrease in aid after a country crosses the IDA threshold. I then run the IV estimates for the full sample. The baseline IV estimates for the full sample confirm the causal estimates of Galiani *et al* (2017) that aid causally increases growth in countries crossing the IDA threshold, on average. The coefficients on aid from these estimates are about twice as large as the OLS estimates, as expected, and they serve as the benchmark estimates for comparison with the IV estimates for aid interacted with the MCC pass-fail dummy variables. I then run the Two-Stage Least Squares (2SLS) model to obtain IV estimates when ODA is interacted with the MCC dummies.

I find that the aid-growth relationship is much stronger for worse-governed and less democratic countries. Across the sub-groups defined by MCC's scorecard, I find that countries that do not pass half of the MCC scorecard indicators, countries that do not pass the democratic rights hard hurdle, and countries that do not pass the MCC scorecard overall tend to be *more* effective in translating aid into economic growth, on average. The only exception is for the control of corruption indicator for which there is no statistically significant difference between passers and failers. On the whole, my results suggest that better-governed and democratic countries perform *worse* than poorly governed and undemocratic countries when it comes to translating aid into economic growth. That is, a performance-based allocation of resources does not translate into better long-term growth outcomes for each dollar of assistance – indeed the opposite is true.

This finding has significant implications for donors because it does not lend support at the macro level for a performance-based approach to aid allocation. The World Bank adjusts up or down their assistance to a country based on its CPIA scores and MCC conditions its decisions of whether to work at all with a country based on whether it passes the scorecard. These resource allocation policies are predicated (at least historically) on the conditional aid-growth literature. Given this new evidence I present that the aid-growth relationship is weaker for better-governed countries, these policies should potentially be reconsidered. Because performance-based allocation effectively limits the range and scope of potentially effective partnerships and given that most of the world's poor live in countries that are poorly governed (Milante *et al* 2016), this finding suggests that donors may be allocating assistance away from countries with the greatest need *and* getting worse returns on investment as a result.

Nonetheless, there was still a strong relationship between aid and growth for the entire sample. The implication is that for this type of country – relatively poor but growing developing countries graduating from grant financing – there is still a strong case that aid is effective in promoting economic growth. However, my findings do not lend support for a performance-based model of allocating aid on the basis of good governance or democracy where this is being justified by improved macro-level outcomes in translating aid into economic growth outcomes. There are other compelling reasons to allocate aid selectively, but the aid-policy-growth relationship does not appear to be one of them.

The first contribution of this chapter is to confirm that aid causes growth, which is in itself a controversial finding. I exploit a plausible instrument to estimate a causal relationship and find that aid is effective in promoting growth among countries crossing the IDA threshold. There is a long literature in development economics related to the question of whether foreign assistance results in economic growth, which has produced mixed and sometimes conflicting results. Galiani *et al* (2017) arguably provided the first plausible causal estimates of aid on growth, and my results confirm their findings and provide additional rigorous evidence that aid is good for growth. Specifically, I confirm the robustness of their finding with eleven new years of data and a modified empirical approach. While this is a compelling finding on its own, it only tests the unconditional question of whether aid effects growth and not the conditional question of whether aid is more effective in promoting growth in better-governed countries. Thus, I also exploit the instrument to test the conditional aid-policy-growth question.

The second contribution is to explore the conditional strand of the aid-growth literature with a plausible instrument, and my findings upset the aid policy orthodoxy and have significant implications. Burnside and Dollar (2000) found that aid has a positive effect on growth *only* in countries with prudent fiscal, monetary, and trade policies. Though the work was criticized due to the fragility of its results, Burnside and Dollar (2000) was highly influential in shifting aid allocation policy towards a more performance-based approach, including MCC's selectivity model.[27] Jia and Williamson (2019) were not able to replicate the Burnside and Dollar (2000) results, and their findings suggest that aid conditional on policy performance does not promote greater economic growth. Jia and Williamson (2019) explored a range of potential instrumental variables, including the IDA threshold, but they did not have a sufficient sample size to exploit the IDA threshold instrument. This is where my research fills a gap in the literature.

That is, the academic literature is missing research that revisits Burnside and Dollar's (2000) finding with updated data and a plausible instrument. While Galiani *et al* (2017)'s methodology is compelling, they only test the unconditional question of whether aid effects growth and *not* the conditional question of whether aid is *more* effective in promoting growth in better-governed countries. Jia and Williamson (2019) thoroughly investigate the aid-policy-growth question, but they are not able to exploit the IDA threshold instrument. With the benefit of several additional years and a slightly modified approach to maximize my sample size, I am able to exploit the IDA threshold instrument to test the conditional question of whether aid goes further in better-governed and more democratic countries. While I find that aid does indeed cause growth, this relationship is actually *worse* for better-governed and more

[27] For example, the World Bank instituted "*ex post* conditionality" that rewarded countries for good policy performance as opposed to relying on credible threats to withdraw funding if *ex ante* policy commitments were not upheld. The findings of Burnside and Dollar (2000) were also at least partially the motivation for MCC's country selectivity model, and these ideas have now been fully codified into country scorecards of policy performance and an annual selection process to select countries as eligible to receive grant assistance. Hayes-Birchler and Staats (2014) provide a description of MCC's process and lessons learned.

democratic countries. These findings are important for donors, particularly since country selectively in allocating aid is often justified on the basis of the conditional aid-growth relationship.

These findings have important implications for donors who may want to reconsider performance-based allocation policies that limit aid to countries with the greatest need. For instance, the IDA allocation model utilizes country performance as a main allocation factor with CPIA ratings comprising 92 percent and previous project performance the other eight percent. I show in the next chapter that good governance matters greatly for the success of individual projects, so the World Bank may want to consider shifting the allocation formula from broader governance towards a greater focus on project performance. In selecting country partners, MCC considers both policy performance and the opportunity to promote economic growth, and I find that these criteria conflict – the greatest opportunity is in worse-governed countries. One policy option would be to get away from the corruption and democratic rights hard hurdles, which would make the scorecard much easier to pass and could be framed as an opportunity to work more with fragile states and emerging democracies. Another policy option would be to exploit the flexibility of the Threshold Program to work with countries that do not (yet) pass the scored. However, MCC points to an important incentive effect of their selective approach, and it is understandable if a given donor does not want to shift allocation criteria towards explicitly favoring poorly governed or non-democratic countries, as this could be perceived as a reward for poor performance. Alternative rationales for a selective approach to resource allocation are not necessarily dependent on the conditional aid-policy-growth literature and may stand up on their own merits, though future research should test these claims more rigorously than they have been to date.

Data

The data I use is at the country-year unit of analysis, and the full dataset covers the period of 1987 through 2021. The sample of countries was determined by availability of the instrument, which I explain below. The sample includes the 54 countries that sustainably crossed the IDA threshold in terms of GNI per capita from 1991 to 2018. These countries are displayed in Table 1 with the year they crossed the IDA threshold.

Table 1. Countries in the sample and year of IDA threshold crossing

Country	Year	Country	Year	Country	Year
Albania	1999	Georgia	2003	Papua New Guinea	2008
Angola	2004	Ghana	2008	Peru	1990
Armenia	2003	Guyana	2000	Philippines	1994
Azerbaijan	2005	Haiti	2011	Samoa	1995
Bangladesh	2015	Honduras	1999	Sao Tome and Principe	2012
Benin	2018	India	2010	Senegal	2007
Bhutan	2003	Indonesia	2003	Solomon Islands	1993
Bolivia	1997	Kenya	2014	Sri Lanka	2003
Bosnia and Herzegovina	1997	Kiribati	1991	Sudan	2007
Cambodia	2017	Laos	2012	Tajikistan	2013
Cameroon	2004	Lesotho	2006	Timor-Leste	2006
China	2000	Mauritania	2006	Turkmenistan	2003
Comoros	2005	Moldova	2005	Ukraine	2003
Republic of Congo	2005	Mongolia	2006	Uzbekistan	2010
Cote d'Ivoire	2012	Myanmar	2016	Vietnam	2010
Djibouti	2007	Nicaragua	1999	Yemen	2012
Egypt	1995	Nigeria	2005	Zambia	2008
Equatorial Guinea	2001	Pakistan	2014	Zimbabwe	2012

Source: World Bank World Development Indicators and Gaiani et al (2017). See Apendix 2.

I obtained and merged data from three different sources: average income and economic growth, foreign aid, and population from the World Bank's World Development Indicators (WDI); economic growth and consumption growth from Penn World Tables (PWT); and pass-fail dummy variables from MCC's historical country scorecards.

My primary data source is the WDI from the World Bank. These data include the outcomes of interest, growth in Gross Domestic Product (GDP) and GNI by country on an annual basis. In addition, I employ GDP growth from PWT as well as consumption growth as dependent variables in my baseline OLS estimates as well as robustness checks in Appendix 4. The primary explanatory variable also comes from WDI, Official Development Assistance as a proportion of GNI (ODA/GNI) by country on an annual basis. These data are originally collected the Organisation for Economic Co-operation and Development (OECD). My control variable, total population, is also from WDI. The population control variable is by country on an annual basis, and I use its natural log.

For the pass-fail dummy variables, I scraped the MCC website for information from their historical scorecards. I produced time-invariant pass-fail dummy variables for each country that are the same across the entire duration of the sample – that is, a country either passes or fails the indicator in all years of the sample based on their status when they crossed the IDA threshold. I produce four different pass-fail dummy variables based on whether a country is: (1) passing half of the indicators on the scorecard; (2) passing the control of corruption indicator (also known as the corruption hard hurdle); (3) passing either the political rights or civil liberties indicators (also known as the democratic rights hard hurdle); and (4) passing the MCC scorecard overall, which entails passing all of the previous three criteria (passing half of the indicators, including the corruption and democracy hard hurdles). This is the same method used by the MCC to determine whether a country can be selected as eligible, though their approach has evolved slightly over time.[28] To determine passing or failing, I examined a country's scorecards for the three years after a country crossed the IDA threshold, if available. See Appendix 2 for a more detailed accounting of my method for determining passing or failing and Appendix 5 for a

[28] For example, MCC used to require that a country pass half of the indicators from three different buckets of indicators – ruling justly, economic freedom, and investing in people. This was simplified in 2012 to only passing half of the indicators overall. See Hayes-Birchler and Staats (2014) for an overview of the scorecard's evolution.

robustness check that shows that omitting countries on the threshold or that are rapidly changing on their scorecard performance does not change my results.

The instrument is a dummy variable that switches from a value of zero to one for the years in which a country has sustainably passed over the IDA threshold in terms of its GNI per capita. The IDA crossing dummy is lagged one year from the explanatory variables (aid, population) and three years from the outcome of interest (growth). I constructed this dummy by comparing the historical GNI per capita to the IDA threshold for each year to determine when the country had sustainably crossed over the threshold. For instance, I did not consider a country to have sustainably crossed the threshold if its GNI per capita crossed the threshold for one or two years only to fall back below the threshold. Alternatively, if a country's GNI per capita was historically above the threshold, I did not count it as crossing the threshold if it dipped below the threshold for one or two years and then crossed again. (This process is described in more detail in Appendix 1.)

I treat the IDA threshold crossing in two ways. First, I code all years subsequent to the crossing as 1 – in effect, this compares all pre-crossing years to all post-crossing years regardless of when the country crossed the threshold. As a result, this could mean there are many pre-crossing years or, more problematically, many post-crossing years. Given the length of the sample, in several countries, this means that there are 25 to 30 years of data after crossing the threshold. Since the large decrease in aid after crossing the threshold likely would only affect the broader economy for a few years, I also construct a second dummy variable which omits country-year observations more than three years after crossing. That is, there are only three values of 1 for the dummy variable, after which I remove the country from the sample. I keep all pre-crossing observations in the sample, however. Although this limits the sample size, it is potentially a more precise comparison of the pre-crossing period to the immediate post-crossing period when the effects of a draw-down in aid are most acute. Both IDA crossing dummies are examined separately in my results below.

Finally, throughout this chapter, I utilize country-years as the unit of observation instead of the three-year periods used by Galiani *et al* (2017) and others. This contributes to greater precision in the timing of the threshold crossing, decrease in aid, and growth effects. For example, if a country passes over the threshold in the first year of a three-year period, aid might drop right away and be captured in years 2 and 3 of that period, which is still the pre-period when the (three-year period) IDA crossing is lagged. Alternatively, a country might cross over the threshold in year 3 of the period, and the decrease in aid may not happen for several years, which means that a decrease in aid may not show up in the data in the period following the crossing, and it would appear that aid did not decrease in response to the country crossing the IDA threshold. Using country-years also allows for more consistency in terms of the timing of the crossing and the subsequent years that are examined for a decrease in aid.

Summary statistics

Table 2 displays the summary statistics for the variables used in my analysis. With 54 countries in my sample that have sustainably crossed the IDA threshold over a 35-year period, there is a maximum of 1,890 observations for any given indicator.

The primary outcome of interest is economic growth. This is measured using both annual GNI growth and annual GDP growth from two different data sources (WDI and PWT). I also examine consumption growth, which is a portion of GDP and a critical measure of development at the household level. I utilize all of these variables in the baseline OLS estimates but not the subsequent estimates, as I employ GDP growth from WDI as the main dependent variable. I conduct a set of robustness checks using the other dependent variables in Appendix 4.

Table 2. Summary statistics

	Observations	Mean	Std Dev	Min	Max	Source
GDP Growth (WDI)$_{c, t}$	1,791	4.05	7.59	-44.90	149.97	WDI
GDP Growth (PWT)$_{c, t}$	1,578	5.13	11.04	-56.61	152.32	PWT
GNI Growth$_{c, t}$	1,163	4.12	6.24	-36.67	49.68	WDI
Consumption Growth (PWT)$_{c, t}$	1,539	4.72	13.49	-65.83	351.16	PWT
ODA/GNI$_{c, t-2}$	1,661	6.97	8.59	-0.29	69.40	WDI
Pass Half of MCC Indicators $_{c}$	1,890	0.65	0.48	0.00	1.00	MCC
MCC Corruption Indicator $_{c}$	1,890	0.56	0.50	0.00	1.00	MCC
MCC Democracy Hurdle $_{c}$	1,890	0.57	0.49	0.00	1.00	MCC
MCC Scorecard $_{c}$	1,890	0.35	0.48	0.00	1.00	MCC
Population (ln)$_{c, t-2}$	1,782	15.96	2.04	11.12	21.07	WDI
IDA Crossing (Three Years) $_{c, t-3}$	1,146	0.14	0.35	0.00	1.00	WGI
IDA Crossing (All Years) $_{c, t-3}$	1,728	0.43	0.50	0.00	1.00	WGI

Note: Italicized indicators are 0-1 dummy variables.
Source: World Bank World Development Indicators; Penn World Tables; World Bank Worldwide Governance Indicators; Freedom House Freedom in the World; and Millennium Challenge Corporation Country Scorecards.

GDP growth from WDI has a mean of 4.05, a standard deviation of 7.59, and a range of -44.9 to 150.0. GDP growth from PWT has a mean of 5.13, a standard deviation of 11.04, and a range of -56.6 to 152.3. While the two sources for GDP growth are similar, PWT has a higher mean, greater standard deviation, and wider range of observations. GNI growth from WDI is more similar to GDP growth from WDI than PWT. It has a mean of 4.12, a standard deviation of 6.24, and a range of -36.7 to 49.7. While the means of these two different outcomes of interest are similar at just over 4.0 percent, GDP growth from WDI has more observations, a much greater range of values, and a larger standard deviation. Consumption growth has a similar mean, 4.72, but a much larger standard deviation, 12.5. This is also reflected in the much wider range from the -65.8 minimum observation up to the 351.1 maximum. GDP growth from

WDI is most useful in terms of statistical power, as it has the most observations with 1,791. GDP growth from PWT has 1,578, GNI growth has the fewest with 1,163, and consumption growth has 1,539.

The main explanatory variable is ODA/GNI on the right-hand side. I obtain 1,661 observations for ODA/GNI, and the variable ranges from -0.3% (as it is *net* ODA) up to nearly 70 percent of GNI. The mean is about 7.2 percent of GNI, and the standard deviation is about 8.7. I employ one control variable, which is the natural log of a country's population with a mean of 15.9.

Turning to the MCC scorecard dummies, the mean of the "pass half" dummy is 0.65. This means that 65 percent of the countries in the sample are passing at least half of the indicators on the MCC scorecard in at least two of the three next years after they crossed the IDA threshold. The mean of the pass-fail dummy for the control of corruption hard hurdle is .56, which means that slightly more than half of the countries in my sample were passing the control of corruption indicator. The pass-fail dummy variable for the democratic rights hard hurdle on the MCC scorecard has a mean of .57, which means that slightly more than half of countries in my sample passed either the political rights or civil liberties indicators (or both) on the MCC scorecard. A country must pass all three of the preceding pass-fail dummy variables to pass the MCC scorecard overall. The mean for the MCC scorecard dummy variable is .35, which means that just over one third of the countries in my sample were passing the MCC scorecard in at least two of the three years after they crossed the IDA threshold. This is in line with a historical observation at MCC that roughly one third of countries consistently pass the scorecard (while another third never pass and the middle third only sometimes pass or transition from passing to failing or vice versa).[29]

[29] Since these dummy variables are time invariant, it is possible that scorecard performance at the time a country crosses the IDA threshold is different from when aid is being implemented and effecting growth, especially for this middle third for which there is rapid improvements or declines in scorecard performance. Appendix 5 shows that my results do not change when several of the countries from the middle third are omitted from the sample.

The instrument of crossing the IDA threshold for all years has a mean of 0.43. This means that there are slightly more pre-crossing observations than post-crossing observations. The average country has nearly 20 country-year observations in the pre- period, and just over 15 years in the post- period. For the IDA threshold dummy variable that is limited to three years, there are fewer observations overall – down from 1,664 to 1,118 – and the mean decreases to 0.14. This means that over 85% of the observations are in the pre-period whereas the three years of post-crossing data make up 14 percent of the sample. The average country has almost 21 years of observations, including the three post-crossing years.

Empirical Strategy

To test the effectiveness of a performance-based allocation policy, I employ a fixed effects model and an IV model to examine whether working with well-governed countries improves the effect of aid on growth at a macro level. I pursue this question by running OLS and IV estimates with interaction terms with dummy variables that effectively split the sample into better and worse governed countries.

My empirical strategy exploits the crossing of the World Bank's International Development Association's (IDA) threshold as an instrument. Following significant academic debate regarding the aid-growth relationship throughout the 2000s, Galiani *et al* (2017) arguably provided the first plausible causal estimates of aid on growth by exploiting the World Bank's IDA income threshold as an instrumental variable using an econometric approach similar to a regression discontinuity design. When a country crosses the IDA threshold in terms of its GNI per capita, this should have no immediate economic implications. Yet when countries pass this arbitrary threshold, this typically initiates the graduation process from IDA eligibility, thereby reducing the country's access to grants and highly concessional lending. Because other donors often follow suit, net Official Development Assistance as a proportion of Gross National Income (ODA/GNI) tends to dramatically decrease.

For the instrument to be relevant and valid, it will need to have a sufficiently strong first stage and maintain the exclusion restriction. If both of these assumptions hold, then I can reasonably claim that my IV estimates of aid on growth are causal. For the former, I can examine the first stage F-Statistic – anything above 10 is generally considered to be a sufficiently strong first stage though the weak instrument critical values for my relatively small sample are about seven for the 10% maximal IV size (Stock and Yogo 2005). (For more details, see Appendix 3.) For the exclusion restriction to hold, I must assume that the IDA threshold is arbitrary and does not impact growth in any way except through its impact on aid allocations. This is plausible as only the World Bank explicitly uses this threshold for decisions related to investment decisions, and it has no other economic meaning (Kenny 2014). While other donors also start to draw down grant assistance in concert with the World Bank, this is not done in a systematic or automatic way and other capital flows outside of grant assistance are not influenced.

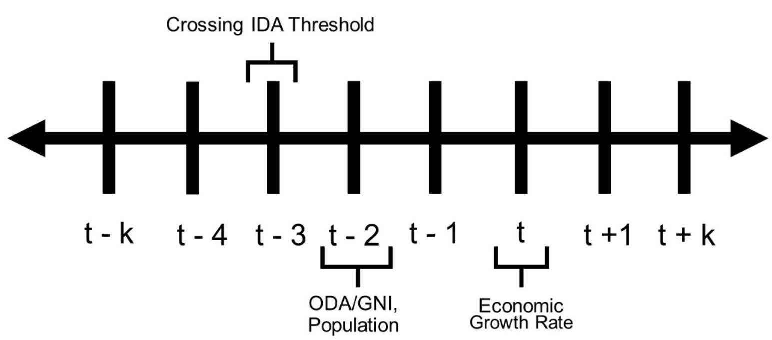
Galiani *et al* (2017) convincingly lay out the case for crossing the IDA threshold as a relevant and valid instrument. The threshold does not necessarily lead immediately to a drop in aid; rather, the graduation process is triggered once a country crosses the threshold, and it often takes several years for the aid reductions to materialize. IDA reductions also serve as a signal to other donors who also start to draw down aid – thus the IDA crossing is amplified beyond just IDA allocations to foreign aid more generally. Galiani *et al* (2017) show that crossing the threshold is typically associated with sizeable reductions in grant assistance, and I confirm this. More generally, there is no reason to believe that crossing an arbitrary threshold that is revised annually and is lagged by more than six months would impact growth except through the assistance channel.[30]

One potential threat to the exclusion restriction would be the behavior of private sector investments in response to these dynamics, and this could influence growth in either direction. Assuming that aid crowds in private capital, investors might draw down their investments in response to lower aid. While it could be argued that this is simply a channel for aid's impact on growth, there is only weak evidence that aid leverages private capital, particularly in lower-income countries (Attridge and Engels 2019). To draw this story out, a critical mass of investors would have to be tracking whether a country crosses the IDA threshold (based on the lagged data) then draw down investment contemporaneously as aid declines, which would then have simultaneous impacts on growth. Alternatively, investors might see opportunities for investments as aid declines to the extent that there are high-return investments that might be profitable to private investors. This also seems far-fetched given the timing issue and the fact that donors tend (and strive) to finance public goods that the private sector would not otherwise finance, e.g., see the World Bank's "cascade framework" (Cordella 2018). One could devise other stories to challenge the exclusion restriction, such as that a drop in aid might worsen the enabling environment and discourage aid, but these stories tend to rely on some combination of heroic assumptions: aid is highly visible and predictable (which it usually is not); aid is extremely impactful at a systemic level (despite its relatively small volume); and the timing of private investor responses line up with both the lagged aid explanatory variable and the growth outcomes.[31]

[30] The World Bank releases it GNI per capita data, analytical classifications, and operational lending categories (that include the IDA threshold) on July 1st each year for the preceding calendar year.
[31] Areas for further research could be related to the effect on private investment of (arguably) exogenous changes in aid flows, such as crossing the IDA or IBRD line or being selected for an MCC compact.

By exploiting this instrument, I can provide rigorous, empirical evidence that aid is good for growth – or in these cases, less aid is bad for growth – and then explore whether this relationship is different for better-governed or more democratic countries. Galiani *et al* (2017) showed that their naïve Ordinary Least Squares (OLS) estimates are comparable in size to the fixed effects estimates of Clemens *et al* (2012), while their instrumental variables (IV) estimates were approximately twice as large. This substantiated concerns that previous aid-growth estimates were biased downwards due to a combination of factors, such as reverse causality and omitted variable bias, e.g., aid going to countries experiencing conflict or a natural disaster that also negatively affected economic growth. I will take the same approach of producing baseline OLS estimates first, then exploiting the IDA instrument to produce IV estimates. I will then take this one step further by interacting MCC scorecard dummies with ODA/GNI.

Figure 1. Timing of Lagged Variables



To account for the timing of the effects, I lag the instrument and all the explanatory variables. The timing for the variables is graphically displayed in Figure 1. The instrument is lagged three years and ODA/GNI, and the control for population are lagged two years. The corresponding narrative here is that once a country crosses the IDA threshold (in *t-3*), this triggers the graduation process. Though it does not immediately or automatically disqualify them from receiving IDA assistance, the World Bank and other donors start to draw down their budget allocations in the years to follow. Realistically, the dramatic reductions to aid will take at least a year to materialize (*t-2*), and it may take multiple years.

In the years following the IDA crossing, the (lower) aid is disbursed and the control for population is relevant (*t-2*). Finally, these potential impacts on economic growth do not materialize immediately. New skills take a while to be used, new infrastructure stimulates local markets over time, and other investments may take years to see results. Thus, I can reasonably expect to see any effects on growth in the second year after the aid is disbursed – the third year following the crossing at the earliest (*t*).

I examine two different econometric models throughout my analysis, fixed effects and IV, to examine the effect of aid on growth. For each, I also add a series of MCC scorecard dummy interaction terms to test the aid-policy-growth question. I walk through each specification in turn here.

The first specification is a fixed effects model that provides baseline OLS estimates with country and year fixed effects:

$$y_{ct} = \beta_0 + \beta_1 Aid_{ct-2} + \beta_2 Population_{ct-2} + \lambda_t + \alpha_c + \varepsilon_{ct} \qquad (1)$$

Where y_{ct} refers to the annual economic growth for country *c* in year t. Aid_{ct-2} refers to ODA as a proportion of GNI for country *c* in year *t-2*. $Population_{ct-2}$ refers to the control variable of the natural log of population for country *c* in year *t-2*.[32] λ_t and α_c are year and country fixed effects, respectively. All variables extend through *t+k* (with *k* being the end of the sample period) with the same lags. Standard errors are adjusted for clustering at the country level for all specifications.

[32] The population variable controls for the commonly noted phenomenon in aid allocation that large-population countries get less foreign assistance than small population countries (in terms of both aid per capita and relative to its GNI, though not in total volume) due to fixed costs and a variety of other factors (McGillivray 2003).

Equation 1 is run for the full sample and then I also add a set of interaction terms to equation (2) to test the aid-policy-growth relationship:

$$y_{ct} = \beta_0 + \beta_1 Aid_{ct-2} + \beta_2 Aid_{ct-2} * Scorecard_c + \beta_3 Population_{ct-2} + \lambda_t + \alpha_c + \varepsilon_{ct} \quad (2)$$

Specifically, in equation (2) I add an additional term, $\beta_2 Aid_{ct-2} * Scorecard_c$, where the aid term is the same as above and $Scorecard_c$ refers to one of multiple different MCC scorecard dummy variables for governance or democracy that are used throughout this analysis for a country *c*. Note that $Scorecard_c$ is time invariant and only varies across countries, so the level term is perfectly collinear with the country fixed effects (α_c) and cannot be separately estimated in this equation. Whether or not governance or democracy has an independent effect on growth is a separate question that I do not address in this paper.[33] Implicitly, I assume that governance and democracy are sufficiently slow-moving that they can be considered fixed within a country over time (or at least during the three-year period after the IDA crossing for my more limited sample) and are absorbed by the country fixed effects. Equation (1) and (2) provide baseline estimates, although it is well recognized that the allocation of assistance may be endogenous to other factors, such as conflict or natural disaster, so naïve OLS estimates of β_1 may be biased downwards.

I then proceed to IV estimates, which are estimated using two-stage least squares (2SLS). For the basic aid-growth relationship, the first stage is estimated using equation (3):

[33] For more on this question, see Kaufmann *et al* (2000) on the relationship between governance and growth or Acemoglu *et al* (2019) for the relationship between democracy and growth.

$$Aid_{ct-2} = \gamma_0 + \gamma_1 IDA_{ct-3} + \gamma_2 Population_{ct-2} + \lambda_t + \alpha_c + \varepsilon_{ct} \qquad (3)$$

Where $Aid_{c,t-2\ t}$ refers to ODA as a proportion of GNI for country *c* in year *t-2*. IDA_{ct-3} refers to a dummy variable for whether a country *c* has sustainably crossed the IDA threshold in the year *t-3*. The second stage in the IV equation is shown in equation (4):

$$y_{c,t} = \beta_0 + \beta_1 \widehat{Aid}_{ct-2} + \beta_2 Population_{ct-2} + \lambda_t + \alpha_c + \varepsilon_{ct} \qquad (4)$$

As in Equation (1) and (2), y_{ct} refers to economic growth for country *c* in the year *t*. $\widehat{Aid}_{ct-2}$ refers to the predicted aid values from the first stage in Equation (3) for a country *c* in the year *t-2*. $Population_{ct-2}$ refers to the control variable of the natural log of population, and λ_t and α_c are the year and country fixed effects, respectively.

While Galiani *et al* (2017)'s IV approach is compelling, they only test the unconditional question of whether aid causes growth and not the conditional question of whether aid is more effective in promoting growth in better-governed countries. Their instrument could be used to revisit the conditional Burnside and Dollar (2000) question, i.e., whether there is an improvement in the aid-growth relationship for better-governed countries. To do this, I produce IV estimates with the MCC scorecard dummy interaction terms to test whether performance-based factors matter for growth outcomes as a result of foreign assistance. The first stage is estimated using equations (5) and (6):

$$Aid_{ct-2} = \gamma_0 + \gamma_1 IDA_{ct-3} + \gamma_2 IDA_{ct-3} * Scorecard_c + \gamma_3 Population_{ct-2} + \lambda_t + \alpha_c + \varepsilon_{ct} \qquad (5)$$

$$Aid_{ct-2} * Scorecard_c = \gamma_0 + \gamma_1 IDA_{ct-3} + \gamma_2 IDA_{ct-3} * Scorecard_c + \gamma_3 Population_{ct-2} + \lambda_t + \alpha_c + \varepsilon_{ct} \quad (6)$$

In equation (5), I add an additional term to the first stage equation (3), $\gamma_3 IDA_{ct-3} * Scorecard_c$, where IDA_{ct-3} refers to the threshold crossing dummy and $Scorecard_c$ is a dummy variable for the various elements of the MCC scorecard that is time-invariant for country c.[34] If the exclusion restriction holds for equation (4) – i.e., if IDA crossing only affects growth via its effect on aid – then it follows that it should hold in equation (7). In other words, if crossing the IDA threshold is indeed conditionally exogenous to growth, it follows that there should be no reason that IDA would differentially affect growth across different country-types except through the effect of aid given that the IV model employs country fixed effects, which control for the level effect of $Scorecard_c$ on growth. I must run this same specification on aid interacted with the MCC scorecard interaction term as part of the first stage. For that specification in equation (6), the dependent variable is $Aid_{ct-2} * Scorecard_c$, where the aid term is the same as the left hand side in equation (5) and the MCC scorecard dummy is the same as the other interaction term on the right hand side. The rest of the specification is the same as equation (5).

The second stage for the IV model with interaction terms is shown in equation (7):

$$y_{c,t,} = \beta_0 + \beta_1 \widehat{Aid}_{ct-2} + \beta_2 \widehat{Aid_{ct-2} * Scorecard_c} + \beta_3 Population_{ct-2} + \lambda_t + \alpha_c + \varepsilon_{ct} \quad (7)$$

This is the same as equation (4) except that I add an additional term, β_2, which refers to the predicted values of the interaction term in equation (6).

[34] Estimating the IV model using equations (5) and (6) and the two instruments $IDA_{ct-3} * Scorecard_c$ and IDA_{ct-3} enables me to avoid the so-called "forbidden regression" (Angrist and Pischke 2009).

Results

I present my results below. I start with OLS estimates for the full sample then compare those to OLS estimates with the MCC scorecard dummy interaction terms. I then run the first stage estimates to test the strength of the instrument and produce baseline IV estimates. I then produce IV estimates with MCC scorecard interaction terms to determine whether there is a differential effect for better governed or more democratic countries. I conclude with my interpretation of the results.

Baseline Ordinary Least Squares Estimates

First, I estimate the baseline OLS relationship using country and year fixed effects. Table 3 presents the results of estimating equation (1) above. The baseline OLS results are shown for four different outcomes of interest. Column (1) uses WDI data for GDP growth, column (2) uses a different measure of GDP growth from PWT, column (3) uses GNI growth from WDI, and column (4) uses consumption growth from PWT. This is the only table in which I examine all four outcomes of interest – from here forward I will only employ the GDP growth variable from WDI. I conduct robustness checks for my subsequent specifications using these alternate dependent variables in Appendix 4. As Table 3 shows, the coefficients on ODA/GNI are similar across all four specifications, so I chose to use GDP growth from WDI as it has the greatest number of observations. This follows the example of much of the aid-growth literature, though not Burnside and Dollar (2000) who use PWT and whose results are not robust to using WDI data instead (Jia and Williamson 2019).

The OLS estimates in Column 1 of Table 3 suggests that a 1 percentage point increase in the aid to GNI ratio from the average level ODA/GNI at the year of crossing is associated with a 0.16 percentage point increase in GDP growth (using the WDI data), which is statistically significant at the 5% level. Column 2 runs the same regression but with GDP growth rate from PWT as the outcome of interest, and I find that a 1 percentage point increase in ODA/GNI is associated with a .18 percentage point increase in growth, which is minimally statistically significant at the 10% level. Column (3) finds that a one percentage point increase in ODA/GNI results in a .23 percentage point increase in GNI growth, which is statistically significant at the 1% level. Column (4) finds a weaker relationship between aid and consumption with a one percentage point increase in ODA/GNI associated with a .11 percentage point increase in consumption growth, which is minimally statistically significant at the 10% level. These results are roughly in line with the literature using fixed effects models (e.g., Clemens *et al* 2012) of about a 0.2 coefficient for ODA/GNI on GDP growth.

Table 3. OLS Estimates of Growth on Official Development Assistance

	(1)	**(2)**	**(3)**	**(4)**
	GDP Growth (WDI)	**GDP Growth (PWT)**	**GNI Growth (WDI)**	**Consumption Growth (PWT)**
ODA/GNI	0.1624**	0.1823*	0.2289***	0.1061*
	(0.0662)	(0.0988)	(0.0766)	(0.0603)
Population	✓	✓	✓	✓
Country Fixed Effects	✓	✓	✓	✓
Year Fixed Effects	✓	✓	✓	✓
Observations	1,636	1,421	1,103	1,421

Note: OLS regression of growth variable on ODA/GNI, population, and fixed effects.
* p < .1, ** p < .05, *** p < .01.
Source: World Development Indicators and Penn World Tables.

Ordinary Least Squares Estimates with Interaction Terms

I next examine various aid-policy interaction terms in equation (2) utilizing the pass-fail dummies for the MCC scorecard ($Scorecard_c$) for passing half of the indicators (i.e., a proxy for generally good governance), passing the control of corruption indicator, passing one of the two democratic rights indicators, and passing the MCC scorecard overall, which requires passing the three previous hurdles. In general, better-governed and more democratic countries tend to have a less clear relationship in terms of aid's effect on growth. The estimates from the interaction terms with the dummy variables are uniformly negative or insignificant – governance, democracy, and the scorecard as a whole are negative, and corruption is statistically insignificant. In other words, none of the specifications suggest that the aid-growth relationship is stronger in better-governed or democratic countries.

Table 4. OLS Estimates of GDP Growth on ODA and Aid*Policy Interaction Terms

	(1)	(2)	(3)	(4)
	GDP Growth			
ODA/GNI	0.2973**	0.0869	0.3039**	0.2156**
	(0.1129)	(0.0846)	(0.1307)	(0.0960)
ODA/GNI*MCC-Half (dummy)	-0.2545**			
	(0.1150)			
ODA/GNI*MCC-Corruption (dummy)		0.1078		
		(0.1160)		
ODA/GNI*MCC-Democracy (dummy)			-0.2448*	
			(0.1342)	
ODA/GNI*MCC-Scorecard (dummy)				-0.1469
				(0.1011)
Population	✓	✓	✓	✓
Country Fixed Effects	✓	✓	✓	✓
Year Fixed Effects	✓	✓	✓	✓
Observations	1,636	1,636	1,636	1,636

Note: OLS regression of economic growth on ODA/GNI and interaction terms. * p < .1, ** p < .05, *** p < .01.
Source: World Development Indicators, Worldwide Governance Indicators, Freedom in the World, MCC .

In column (1) of Table 4, the ODA/GNI term is positive and statistically significant at the five percent level, while the MCC-Half interaction term is negative and statistically significant at the 5% level. A one percentage point increase in ODA/GNI is associated with a .30 percentage point increase in GDP growth, but in the countries that pass half of the MCC indicators, they experience a .25 percentage points smaller increase in economic growth. This suggests that there is a larger effect of aid on growth in *worse*-governed countries. This is an unexpected result. For both ODA/GNI and the corruption dummy interacted with ODA/GNI, I find no significant results for either term in column (2). However, it is notable that this is the only specification that results in both coefficients with the expected signs, i.e., both positive, which would suggest that aid promotes growth in all countries but goes further in better-governed countries (but does not as they are both statistically insignificant).

For the specification with the democracy hurdle dummy, I again find that less democratic countries grow faster as a result of aid. In column (3), for countries that do not pass the democracy hard hurdle, I find that aid is associated with economic growth – a one percentage point increase in ODA/GNI produces a .30 percentage point increase in GDP growth that is statistically significant at the five percent level. However, the countries that pass the democracy hard hurdle do not have the same relationship – they experience .24 percentage points less GDP growth than undemocratic countries, which is statistically significant at the 10% level. These estimates are nearly identical to column (1) for the dummy that represents passing half of the scorecard indicators. This suggests that less democratic countries translate aid into economic growth whereas democracies do not.

When putting together the pass half, corruption, and democracy pass-fail indicators, only about 35% of the sample pass all three indicators and thus pass the MCC scorecard. While countries failing the scorecard see an association between aid and growth in column (4), the differential relationship for MCC scorecard passers in the interaction term is negative but not statistically significant. A one percentage point increase in ODA/GNI results in a .22 percentage point increase in GDP growth for failing countries, which is statistically significant at the 5% level. The coefficient on the MCC scorecard pass-fail dummy is negative, which is the opposite of what I expect. This is particularly surprising considering that these countries must be reasonably non-corrupt democracies that have generally good governance, at least relative to their peer developing countries.

This initial indication that there is not an increase in the effectiveness of aid in terms of growth for the better-governed countries is troubling for those that hold up country selectivity as a critical enabler of aid effectiveness. Rather, these estimates suggest that it is possible that the relationship between aid and growth may actually be stronger for worse governed and less democratic countries. These results establish the baseline for comparison to the IV estimates, which will provide causal estimates.

Baseline First Stage Estimates

Whether crossing the IDA threshold is a relevant instrument for aid can be tested by looking at the first stage of a 2SLS regression from equation (3). As described in the data section, I have two different dummy variables for crossing the IDA threshold. The first (column 1) only accounts for what happens to aid in the first three years after the country crosses the threshold with the rest of the years in the sample omitted. The second dummy variable (column 2) considers all years left in the sample after the country crosses the threshold. The IDA threshold crossing variable is lagged one year to capture the change in aid that follows a country crossing the IDA threshold. Table 5 shows the results of the first stage regressions for both samples.

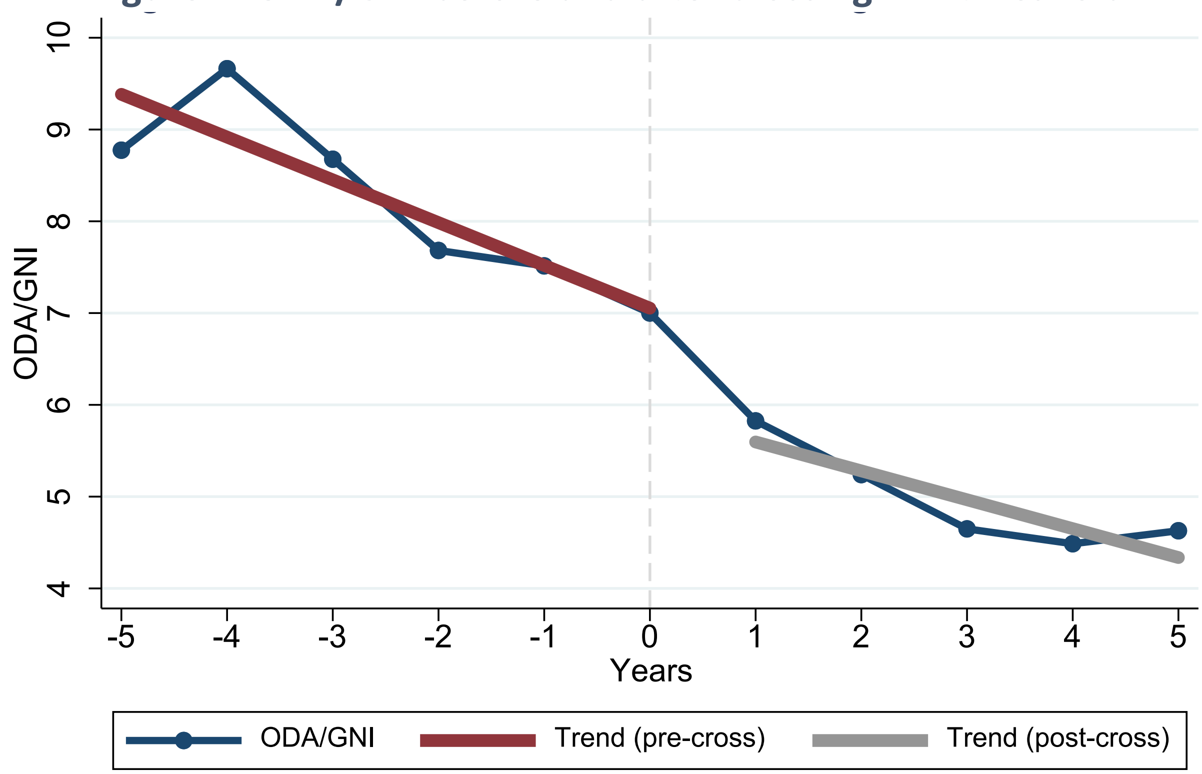
Table 5. First Stage Estimates of ODA/GNI on IDA Threshold Crossing Dummy

	(1)	**(2)**
IDA Crossing (Three Years)	-3.3124***	
	(0.8694)	
IDA Crossing (All Years)		-3.2319***
		(0.9306)
Population	✓	✓
Country Fixed Effects	✓	✓
Year Fixed Effects	✓	✓
F-Stat	14.50	12.10
Observations	1,623	1,040

Note: OLS regression of ODA/GNI on lagged dummy for crossing IDA threshold. * p < .1, ** p < .05, *** p < .01.
Source: World Development Indicators.

For both dummy variables, there is a strong and statistically significant decrease in ODA/GNI after a country crosses the IDA threshold. Column 1 shows that when a country crosses the IDA threshold, it experiences a 3.3 percentage point decrease in ODA/GNI for the three years that follow, on average, that is statistically significant at 1% level. Importantly, the first stage F-statistic is 14.5, which is reasonably strong. Column 2 shows a similar, though slightly weaker, relationship for the entirety of the post-crossing period. When a country crosses the IDA threshold, it receives 3.2 percentage points of GNI less aid for the rest of the sample period, on average. This is also statistically significant at the 1% level, and the F-stat is 12.1.

Figure 2. ODA/GNI before and after crossing IDA threshold



Source: World Bank World Development Indicators.

The decrease in aid is shown graphically in Figure 2. This shows an average of ODA/GNI for all countries in the sample both before crossing the threshold and following it. The timing of the first year of the IDA crossing dummy for all countries has been aligned to t=0, so t=1 is the first post-crossing year in which a country is expected to see large decreases in aid. In the years before crossing the IDA threshold, countries are well above the mean for ODA/GNI of about seven, on average. There is a sharp decrease following crossing the IDA threshold, and the mean drops below six in first year after crossing.

Baseline Instrumental Variables Estimates

My IV estimates for the second stage in equation (4) are shown in Table 6. Column 1 shows that a one percentage point increase in foreign aid causes a .54 percentage point increase in GDP growth, which is statistically significant at the 5% level. This effect is slightly smaller when only looking at the three years following the crossing of the IDA threshold. Column 2 shows that a one percentage point increase in ODA/GNI leads to a .43 increase in GDP growth, which is statistically significant at the 10% level.

Table 6. Instrumental Variable Estimates of GDP Growth on ODA

	(1)	**(2)**
	GDP Growth	
Crossing Duration:	**All Years**	**Three Years**
ODA/GNI	0.5389**	0.4304*
	(0.2117)	(0.2276)
Population	✓	✓
Country Fixed Effects	✓	✓
Year Fixed Effects	✓	✓
Observations	1,598	1,020

Note: 2SLS regression of GDP growth on ODA/GNI instrumented by IDA treshold crossing, population, and fixed effects.
* p < .1, ** p < .05, *** p < .01.
Source: World Development Indicators.

In general, these specifications confirm the findings of Galiani *et al* (2017) that aid causes growth in countries crossing the IDA threshold. This is an important contribution because there is a long literature in development economics related to the question of whether foreign assistance results in economic growth, which has produced mixed and sometimes conflicting results. Galiani *et al* (2017) arguably provided the first plausible causal estimates of aid on growth, and my results confirm their findings and provide rigorous, empirical evidence that aid is good for growth. I confirm the robustness of their findings with eleven new years of data and a slightly modified empirical approach.

These estimates will serve as useful benchmarks when looking at the IV estimates with interaction terms included to examine the effect of governance. It is notable that the coefficients for ODA/GNI are roughly twice as large the OLS estimates, which is very similar to the Galiani *et al* (2017) findings when comparing their OLS and IV estimates. While this is a compelling finding on its own, this only tests the unconditional question of whether aid effects growth and not the conditional question of whether aid is more effective in promoting growth in better-governed countries. Thus, I also exploit the instrument to test the conditional aid-policy-growth question.

IV Estimates with Interaction Terms

The results of my 2SLS specifications with MCC scorecard dummy interaction terms are presented in Table 7. The joint first stage F Statistic for equations (5) and (6) is reported in the second to last line of the table. In seven of the eight specifications, the F-stat is above the standard rule of ten that suggests the first stage is sufficiently strong. For column (7), the F-stat falls below ten at 8.6, but this still exceeds the Stock-Yogo (2005) weak ID test critical value of 7.03 at the 10% maximal IV size. This suggests that I do not have a weak instrument – and if the exclusion restriction holds – I can interpret my IV estimates in Table 7 as causal. The full first stage and reduced form results are shown in Table A1 in Appendix 3.

Table 7. 2SLS Estimates of GDP Growth on ODA and Aid*Policy Interaction Terms

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Instrument:	**IDA Crossing (All Years)**				**IDA Crossing (Three Years)**			
ODA/GNI	1.0226***	0.4829	0.9902***	0.7060***	0.9425**	0.3286	0.7945**	0.5606*
	(0.3430)	(0.3424)	(0.3001)	(0.2644)	(0.4357)	(0.4508)	(0.3689)	(0.2997)
ODA/GNI*MCC-Half	-0.7170**				-0.9078**			
	(0.2861)				(0.4428)			
ODA/GNI*MCC-Corruption		0.0772				0.1415		
		(0.3160)				(0.4212)		
ODA/GNI*MCC-Democracy			-0.7221***				-0.6677*	
			(0.2354)				(0.3638)	
ODA/GNI*MCC-Scorecard				-0.3833*				-0.3910
				(0.2151)				(0.3131)
Population	✓	✓	✓	✓	✓	✓	✓	✓
Country Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
Year Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
First Stage F-Statistic	18	16.6	19	18.7	10.9	8.6	12.3	11.6
Observations	1,598	1,598	1,598	1,598	1,020	1,020	1,020	1,020

Note: 2SLS regression of GDP growth on ODA/GNI and interaction terms instrumented by IDA threshold crossing, populatic and fixed effects. * p < .1, ** p < .05, *** p < .01.
Source: World Development Indicators, MCC.

I first examine the specifications that include the interactions terms with the dummy for passing half of the MCC scorecard indicators in columns (1) and (5). In the first column that includes the entire post-crossing period, the coefficient for ODA/GNI is 1.02 and is statistically significant at the 1% level, and the interaction term is -.72 and statistically significant at the 5% level. This means that every one percentage point increase in ODA/GNI countries in the sample caused a 1.02 percentage point increase in GDP growth in countries that did not pass half of the MCC scorecard indicators, whereas in countries that did pass half, aid only caused about a .31 percentage point increase in GDP growth. In column (5) that includes only three post-crossing country-years, the estimates are similar, but the interaction term is even larger in absolute terms (i.e., more negative). The coefficient for ODA/GNI is .94 (for countries that did not pass half), while the interaction term between ODA/GNI and the dummy for passing half has a coefficient of -.91. Both are statistically significant at the 5% level. This implies that there is a strong aid-growth relationship for poorly governed countries and essentially no effect of aid on growth for better-governed countries.

I turn now to the specifications in columns (2) and (6) that include the interaction term for passing the control of corruption indicator. Both the ODA/GNI coefficient and the interaction term are positive in both specifications (though relatively small), which is what I expect across all specifications, but these estimates are not statistically significant at conventional levels. This suggests that corruption does not have a significant effect, on average, in either direction.

In columns (3) and (7) that include the interaction term with the democracy hard hurdle dummy, I find results similar to the "pass half" interaction term specifications. In column (3) for the full sample, the coefficient on ODA/GNI is .99, and the interaction term is -.72. Both are statistically significant at the 1% level. In column (7) including just three post-crossing country years, the coefficient on ODA/GNI is .79, which is statistically significant at the 5% level. The coefficient on the democracy interaction term is -.67, which is statistically significant at the 10% level. This suggests that the aid-growth relationship is much stronger for non-democracies than democracies.

Finally, columns (4) and (8) show the specifications with the interaction term for passing the MCC scorecard overall. Recall that only about one third of the sample passes – the best-governed democracies in my sample. In column (4) with all country-year observations, the coefficient on ODA/GNI is .71, which is statistically significant at the 1% level. The scorecard interaction term is -.38 and statistically significant at the 10% level. The relationship for the specification with three post-crossing years in column (8) is weaker, however. The coefficient on ODA/GNI is .56 and statistically significant at the 10% level. The scorecard interaction term is negative but not statistically significant. These findings definitively show that countries passing the MCC scorecard do not more effectively translate aid into economic growth, on average. Rather, this provides some weak evidence that countries failing the MCC scorecard actually do better.

In sum, all specifications except (2) and (6) show a positive and significant aid-growth relationship in the failing countries (i.e., the ODA/GNI coefficient) but a negative interaction term, which suggests that aid causes *less* economic growth in better-governed countries and democracies. The specifications with the "passing half" interaction terms in columns (1) and (5) are the most pronounced finding, but this generalization applies to all specifications except corruption in columns (2) and (6). Appendix 5 shows that these results are robust to using a more limited sample of countries that are clearly passing or failing each scorecard criteria by omitting countries on the median or threshold, countries with limited historical data, and/or countries that are rapidly changing in terms of their scorecard performance during the three-year period of their IDA crossing.

Taken together, these findings contradict the belief that better-governed or more democratic countries are performing better than worse governed or undemocratic countries at the macro level when it comes to translating foreign aid into economic growth. In all of my specifications, there was no significant positive difference for better-governed countries, at best, and in most specifications the worse-governed or undemocratic countries actually had a stronger aid-growth relationship.

Summary of Results

My baseline OLS estimates from my fixed effects model were in line with the existing literature, but there was a more mixed picture when including the aid*policy interaction terms. In deploying the fixed effects OLS models with interaction terms, I find that better-governed and more democratic countries had a weaker relationship in terms of aid's effect on growth. These OLS estimates suggested that it is possible that the relationship between aid and growth may actually be stronger for worse-governed and less democratic countries, which would put in question the idea of performance-based country selectivity. My 2SLS results confirm that aid causes growth. This updates the Galiani *et al* (2017) finding by increasing the number of countries that have crossed the IDA threshold by more than half and adding eleven new years of data. My IV results find that aid causes growth. While this is a compelling finding on its own, this only tests the unconditional question of whether aid causes growth and not the conditional question of whether aid is more effective in better-governed countries.

Thus, I also exploit the instrument to revisit the conditional aid-policy-growth question with governance-related interaction terms borrowing from the MCC scorecard's pass-fail approach. In general, I found a pattern that better-governed and more democratic countries had a *weaker* aid-growth relationship than worse governed and undemocratic countries. This suggests that there is a significant difference between better and worse governed countries or democratic and undemocratic countries in this sample, but it is in the oppositive direction than was previously understood. That is, aid goes further in terms of economic growth in poorly governed and undemocratic countries.

Taken together, the implication is that for this type of country – poor but growing developing countries graduating from grant financing – there is a strong case for the effective use of aid in general. However, my findings do not support a performance-based approach to allocating aid on the basis of good governance or democracy when this is being justified by better macro-level outcomes in translating aid into economic growth. There are other compelling reasons to allocate aid selectively, as discussed below, but the aid-policy-growth relationship does not appear to be one of them.

Policy implications and discussion

This chapter examined whether foreign aid is more effective in promoting growth in better-governed or more democratic countries. In examining this question, I exploited a plausible instrument in the IDA threshold and confirmed that the instrument predicts a significant decrease in aid after a country passes the IDA threshold and that there is a causal relationship between foreign aid and economic growth. Second, I examined whether there is an improvement in the aid-growth relationship for better-governed countries, and I found that better governed countries exhibit a negative and statistically significant relationship between aid and growth compared to worse-governed and undemocratic countries. This provides evidence that there is not a strong relationship between better governance and stronger aid-growth outcomes. These findings have important implications for donors.

Importantly, these findings do not support performance-based models of aid allocation at the macro level. From the donor perspective, this is a troubling finding – in short, allocating on the basis of policy performance translates into *worse* growth outcomes. While the World Bank adjusts their assistance to a country up or down based on its governance (as measured primarily by CPIA), MCC conditions its decisions of whether to work *at all* with a country based on whether it passes its country scorecard. Both the MCC and World Bank allocation models are predicated (at least historically) on the conditional aid-growth literature. Given my evidence that aid is more effective in *worse*-governed countries, these allocation models should be reconsidered as they effectively limit the range and scope of potentially effective partnerships and exclude from consideration many countries with the greatest need and largest number of poor.

Given my finding that the aid-growth relationship is actually greater in worse-governed countries, a heavy emphasis on governance as an allocation factor does not make sense. (The World Bank is famously and controversially agnostic about democracy, so that is not a relevant consideration for them.) If the World Bank does not feel comfortable favoring worse-governed countries,[35] it could also consider shifting the allocation formula towards a greater focus on need, e.g., the revised IDA allocation formulate proposed by DFID (Dercon and Lea 2016). The implications of shifting the IDA allocation model towards a greater weight on need (as proxied by GNI per capita) is explored in the final chapter.

As it stands, however, the IDA allocation model favors better-governed countries. Country performance ratings (CPR) are the main factor in determining resource allocations, and CPIA ratings comprise 92 percent of the CPR score with previous portfolio performance being the other eight percent.[36] As a result of my findings, the World Bank could ask whether it should more heavily weight historical project performance instead. A preference for broad, country-level development progress over project performance is reflected in the allocation criteria that are overwhelmingly skewed towards country-level policy performance over *ex post* project outcomes; however, my findings suggest that the World Bank should consider assigning more weight to countries that have better historical project outcomes given that the relationship between good governance and economic growth is not as clear-cut as previously assumed. A shift towards emphasizing previous project performance would accomplish two objectives: first, it would shift the allocation criteria away from good governance that does not achieve improvements in the aid-growth relationship; and second, it would reward good governance in a more targeted way by allocating resources to countries that leverage their governance systems to improve the effectiveness of aid (as opposed to the current, poorly targeted reward for good governance).

[35] This is a more complex factor than it appears on the face. While it may be tempting for a policy maker to reject out of hand the idea of rewarding poorly governed countries due to a potential perverse incentive, it could also be argued that governance could also be considered as part of country need. This consideration is particularly salient in the policy discussion around fragile states where it is unclear whether resources should be allocated to the most fragile countries or emerging reformers (McGillivray 2006).

[36] The World Bank also includes population size to scale resources appropriately and average income to incorporate a measure of need. These issues are explored further in the final chapter.

In the case of MCC, the policy implications are more challenging because the agency's identity is so closely linked to its data-driven country selectivity model. MCC considers three factors in choosing countries: policy performance (as measured by the country scorecard and other supplemental information); the opportunity to promote economic growth and poverty reduction; and the availability of funding. Setting aside funding, I have found that the first two factors conflict with each other. I have found that the greatest opportunity to translate aid into economic growth at the macro level is in worse-governed countries, and this is directly in conflict with MCC's scorecard criteria. To pass the MCC scorecard, a country must pass half of the 20 indicators, including control of corruption and one of two democratic rights indicators; however, my results show that countries passing half of the indicators and/or passing the democratic rights hard hurdle did not experience as much growth as a result of aid. This was also the case with passing the scorecard overall, though that relationship was not as statistically significant. The only hard hurdle that potentially improved the aid-growth relationship was control of corruption, but that was not statistically significant.

To address my findings, one policy option for MCC would be to get away from the democratic rights hard hurdle. This would be politically difficult, but it could be framed as an opportunity to work with emerging democracies that are not yet passing the hurdle. Alternatively, such a change could be packaged as part of a broader overhaul of the scorecard drawing on new evidence of what matters most for the aid-growth relationship. This would recognize my findings as problematic for the current set of scorecard indicators and look to establish a different set of indicators that are more fit for the purpose of allocating assistance in pursuit of economic growth and poverty reduction. Further research could isolate governance variables, if any, that are associated with a stronger aid-growth relationship and include them on the scorecard or elevate them as a hard hurdle if they are already included.

Another policy option would be to exploit the flexibility of the Threshold Program (THP) under MCC's Section 616 authority of the Millennium Challenge Act. In short, THPs are smaller programs focused on policy and institutional reforms with countries not yet passing the MCC scorecard to test the government's commitment and the potential for a larger compact investment. The statute does not specify how close a country must be to passing the scorecard to be selected, so MCC has significant flexibility to work with countries that are further from passing the scorecard. This would open up the possibility of working with countries that do not pass half of the indicators or the scorecard overall, but it also carries significant perceived reputational risk in diluting its brand as being selective on the basis of good governance. While it is easy to downplay the importance of this brand given my findings, there are other benefits that are commonly associated with this selectivity model.

Beyond the aid-growth relationship, MCC points to an important incentive effect – the so-called "MCC Effect"– of inducing policy reforms (*ex ante*) in response to its country selectivity process, though the evidence is primarily anecdotal. [37] Even if aid does not go further in better governed countries, the prospect of a large investment may induce policy reform that could benefit growth indirectly over the longer term. MCC's selection of a given country may also serve as a signal of good governance to the private sector that could help to attract private sector investment, though the evidence for this claim is even more shaky and based exclusively on anecdotes. These incentive effects are plausible due to the discrete nature of MCC's investments and the highly transparent and visible selection process that only happens once per year. That is, MCC publishes its country scorecards each year and then its Board selects new country partners from a candidate pool that is greatly constrained by the country

[37] See https://www.mcc.gov/who-we-select/mcc-effect for MCC's description of this phenomenon.

scorecards. If countries were chosen on a rolling basis or they were like other donors with (perpetual) ongoing country relationships, these incentive effects would likely be greatly diminished.

Generalizing beyond MCC, it is understandable if a given donor does not want to shift allocations towards poorly governed or non-democratic countries, as this could be perceived as a reward for poor policy performance. These rationales for a selective approach to resource allocation are not necessarily dependent on the conditional aid-policy-growth literature and may stand up on their own merits, though future research should test these claims more rigorously than they have been to date. For the World Bank, however, they engage with nearly every developing country, so it is more difficult for them to make this argument. They also have credibility problems when it comes to reducing aid based on a deterioration in policy performance (unlike MCC's sometimes harsh Suspension and Termination Policy).

Finally, I share five issues that warrant further research. First, there is a question related to the external validity of these findings, though this is a dwindling concern as there are not many IDA-eligible countries left.[38] A major limitation of this instrument is that it inherently limits the sample to one particular stage of development. It only examines the effects of drawing down aid at the level of average income associated with the IDA threshold and not the aid-growth relationship at other levels of development (for richer or poorer countries). It is possible that the aid-growth relationship in those circumstances might be different. Future research might investigate the robustness of my results by testing whether the causal relationship found here also applies to circumstances beyond the drawdown in aid precipitated by crossing the IDA threshold, e.g., the graduation from UMIC to High-Income Country after which donors are incentivized to draw down aid as it no longer counts as official ODA.

[38] While this is a reasonable concern, it is important to keep in mind that my sample size is 54 countries, which is a significant proportion of all developing countries. In 2019, there were fewer than 80 LICs and LMICs and fewer than 30 countries still below the IDA threshold.

Second, the aid-growth relationship may have complex differences in its dynamics in autocracies versus democracies that are not captured here. Though it is also a long and contested literature, research on the democracy-growth relationship finds that autocracies grow more slowly on average over the long run (Acemoglu *et al*, 2019), but they also experience more volatility in growth (Easterly, 2011) – that is, their highs are higher, but their lows are lower. Seeing as how my sample is limited to those countries crossing the IDA threshold – which necessarily includes strong economic performers and excludes stagnant non-performers – it is possible that I am picking up the effect of a sub-set of high-growth autocracies in my sample (and not the autocratic non-performers) and comparing them to a sub-set of well-governed democracies with slow-and-steady growth (recognizing that almost all rich countries that graduated prior to 1987 are indeed democracies). There is evidence that poorly-governed countries and autocracies are more susceptible to downside risks (Imam and Temple 2023), so a rapid drawdown of assistance could make the original aid-growth relationship appear to be larger if autocracies are less resilient to financial shocks. Future research might consider how the effectiveness of aid interacts with greater resilience to shocks and if and how this is enabled by good governance and democracy.

Third, there is evidence of a tendency by autocracies to deliberately mis-report official statistics to exaggerate growth (Martinez, 2022). However, this misreporting mainly happens after a country graduates from IDA assistance and would work in the opposite direction of what is observed here, i.e., exaggerated growth data post-crossing would translate into a smaller aid-growth relationship when compared to (less exaggerated data from) democracies. Per Jerven (2013), it is important for future research to question official statistics and search for alternative data that could test the robustness of my results, such as using nighttime lights as the dependent variable. The various different dependent variables from different sources investigated in Appendix 4 is a first step in this direction.

Fourth, there are differences between the sub-samples in the size of the decreases in aid after crossing the IDA threshold, which is shown in Appendix 3. Better-governed countries and democracies saw larger decreases in aid, though this difference is not statistically significant at conventional levels. It is well established through a literature on absorptive capacity that there are diminishing returns to aid, particularly at high levels of aid to GDP (Carter 2014). This is essentially that argument in reverse – two to three percentage point decreases in ODA as a proportion of GNI may have significant negative effects, whereas decreases beyond that do not have a marginal effect. This would show up as poorly governed countries having a larger decrease in economic growth as a result of the drawdown in assistance on average, whereas better-governed democracies would have a smaller decrease in growth *per percentage point decrease* in ODA/GNI. This runs counter to the absorptive capacity story and presents a puzzle. Future research might examine my finding through the lens of the absorptive capacity research (e.g., Feeny and De Silva 2012), which might shed light on which types of projects are likely to be drawn down first and their relative effectiveness.

Finally, the aid-growth relationship runs through many channels, which is part of the reason why it has been so difficult to pin down causal estimates. There are a multitude of factors that go into economic growth, and foreign aid is often a drop in the bucket compared to other development finance resources, e.g., private sector investment, remittances, and government spending. This does not necessarily mean that aid is not effective in promoting economic activity – just that it is extremely difficult to identify precise effects causally. This is also a potential explanation for the "micro-macro paradox," which observes that donors often claim project-level success, but it is hard to show that these outcomes add up to a significant impact on economic growth at the macro level (Mosley 1986). Put differently, the aid-growth relationship is the result of a long and complex causal chain, and a lot of factors go into (aggregate) aid's success or failure. There is some evidence (e.g., Denizer *et al* 2013) that suggests that cross-country studies like this miss the point – while there are some differences across countries, there is much more variation in the impact of aid programs *within* countries based on factors like project manager quality. This is a finding that I will explore further in the next chapter and for which I will have to weigh those micro-level findings against the macro-level findings presented in this chapter.

In conclusion, my findings are potentially problematic for selective donors employing performance-based approaches to resource allocation. I find that the aid-growth relationship is much stronger for worse-governed and less democratic countries, and this presents a serious challenge to the donor orthodoxy. On the other hand, it eases the tension with needs-based approaches and other allocation models that favor fragile states of other vulnerable countries that would receive relatively less assistance from a traditional performance-based model. Given this new evidence, there is an opportunity to consider decades-old orthodoxies concerning the most effective use of assistance. Donors should think hard about what they are trying to accomplish with their allocation models and let the evidence guide them to the most appropriate criteria.

Appendix 1. Sample selection

For my instrumental variable, I construct a dummy variable for when a country sustainably crosses the IDA threshold. The IDA threshold is adjusted over time for inflation. It starts out at $580 in 1987 and is currently at $1,255 for 2021 (i.e., the Fiscal Year 2023 IDA threshold). I compared the GNI per capita of a country to the IDA threshold for each year of the sample, and this indicates the timing of the threshold crossing. That is, I determined that a country crossed the IDA threshold when they went from a GNI per capita below the IDA threshold in one year to having a GNI per capita in the next year. Given the variability in both countries' GNI per capita as well as the moving IDA threshold, I decided that a country must stay above the IDA threshold for three consecutive years for it to be counted as a sustainable crossing. This is because many countries go back and forth below and above the line or cross the line for just one or two years before falling below it again for an extended period of time. Presumably, this would not immediately trigger the IDA graduation process as the country either experienced a temporary growth spurt that dissipated or experienced a negative growth episode just after passing the threshold. These situations would not likely trigger a decrease in assistance, so I do not consider them to be a sustainable passing of the threshold.

Furthermore, when examining the updated data, many countries were off by a year or two with respect to Galiani *et al*'s (2017) crossing year. Presumably this was due to later revisions of GNI per capita data. Because aid allocation decisions are made with the information available at that time, I chose to use Galiani *et al*'s (2017) crossing year in almost every case where there was a discrepancy. This is at least partially because these decisions are not straightforward. For instance, Comoros started out above the threshold at the beginning of my dataset (1987) but then fall below the threshold in 1995. Then, a decade later it "graduated" again in 2005.

Given the criticisms of Burnside and Dollar (2000) that show that sample selection can be critically important for the robustness of results, I aim to be as transparent as possible in explaining why I chose which countries to include and when they crossed the IDA threshold. Therefore, I give a brief description below of any country for which it is not obvious when they crossed the IDA threshold sustainably and permanently when comparing their GNI per capita to the IDA threshold. That is, I do not explain countries that start with a GNI per capita below the threshold, cross the threshold at some point in the sample period (1987-2021), and do not fall back below the IDA threshold by the end of the sample. The year of crossing must also align with Galiani *et al* (2017). In all other cases, I explain how I determined whether to include the country and the year it crossed the IDA threshold sustainably.

- **Albania (1999):** Albania started out above the IDA threshold in 1987 but dropped below it from 1990 through 1998. Thus, I determined that Albania sustainably crossed the IDA threshold in 1999.
- **Angola (2004):** Angola started out above the IDA threshold in 1987 but dropped below it for one year in 1988. They then crossed the threshold again in 1989 before dropping below it again in 1992 for over a decade. Angola sustainably crossed the IDA threshold in 2004.
- **Benin (2018):** Benin crossed the IDA threshold in 2013 but fell back below it in 2015. It sustainably crossed in 2018.
- **Bolivia (1997):** Bolivia started out above the IDA threshold in 1987 but fell below it in 1990 before crossing the threshold in 1997. It then fell below the threshold in 2003 before again crossing it in 2006. Because Bolivia was only below the threshold for two years, I chose the first crossing, 1997.
- **Cameroon (2004):** Cameroon started out above the IDA threshold in 1987 but fell below it in 1994. It then sustainably crossed the IDA threshold in 2004.
- **Comoros (2005):** Comoros started above the threshold in 1987 but then fell below it in 1995. It then sustainably graduated in 2005.

- **Republic of Congo (2005):** The Republic of Congo started above the IDA threshold but fell below it in 1994 before crossing again in 2005.
- **Cote d'Ivoire (2012):** Cote d'Ivoire started the sample above the threshold but fell below it in 1991. They crossed the threshold in 2009 but fell below it in 2011 before sustainably passing in 2012.
- **Djibouti (2007):** There are currently no GNI per capita data in WDI for Djibouti, so I used the same year as Galiani *et al* (2017), 2007.
- **Egypt (1995):** Egypt started the sample above the IDA threshold but fell below it in 1991 before sustainably crossing the threshold in 1995.
- **Honduras (1999):** Honduras started above the threshold but fell below it in 1995. They sustainably crossed the IDA threshold in 1999.
- **Indonesia (2003):** Indonesia crossed the IDA threshold in 1995 but fell back below it in 1998. They sustainably crossed the threshold in 2003.
- **Iraq (excluded):** Iraq started out well above the IDA threshold but then dropped below it in 1991 due to the first Iraq War. They then crossed the threshold in 1998 before falling below it again in 2003 due to the second Iraq War. They then crossed the threshold again in 2004. Because of the extreme fluctuations in GNI per capita due to the conflict and the country's level of general development that is well above the IDA threshold, I do not include this case.
- **Kyrgyz Republic (excluded**): The Kyrgyz Republic crossed the IDA threshold in 2014 but fell below it the next year in 2015. It then crossed the IDA threshold again in 2018 but fell below it in 2020. Because it did not sustainably cross the threshold for three years, I did not include it in the sample.
- **Mauritania (2006):** Mauritania started out above the IDA threshold but fell below it in 1990. It sustainably crossed the threshold in 2006.
- **Mongolia (2006):** Mongolia started out above the IDA threshold but fell below it in 1993. They sustainably passed the IDA threshold starting in 2006.

- **Myanmar (2016):** Myanmar crossed the IDA threshold in 2016, though it fell below it in 2021. Because the country was above the threshold for five years, I opted to include it even though it fell back below the threshold in the latest year of data.
- **Nepal (excluded):** Nepal only crossed the IDA threshold for one year in 2019 before it fell back below it in 2020 and 2021, so I decided not to include it.
- **Nicaragua (1999):** Nicaragua started out above the IDA threshold but fell below it in 1989. It sustainably crossed the IDA threshold starting in 1999.
- **Nigeria (2005):** Nigeria started out above the IDA threshold before falling below it in 1989. It sustainably crossed the threshold starting in 2005.
- **Papua New Guinea (2008):** Papua New Guinea started out above the IDA threshold before falling below it in 1997. It sustainably crossed the IDA threshold starting in 2008.
- **Peru (1990):** Peru was above the IDA threshold for the entire sample in my data, but Galiani *et al* (2017) found that they crossed in 1990, so I used that year instead of omitting it from my sample.
- **Philippines (1994):** The Philippines was above the threshold for the entire period in my data, but Galiani *et al* (2017) found they crossed in 1990, so I used that instead of omitting it from my sample.
- **Samoa (1995):** I did not have GNI data for 1992 through 2001 and in the first year for which I have data, 2002, Samoa was well above the IDA threshold. Therefore, I rely on Galiani *et al* (2017) who determined that Samoa crossed the threshold in 1995.
- **Senegal (2007):** Senegal started out above the IDA threshold but fell below it in 1994. They sustainably crossed the threshold starting in 2007.
- **Solomon Islands (1993):** The Solomon Islands started out just above the IDA threshold in 1987 then fell below it in 1988. They crossed the threshold in 1993 before again falling below it in 2002. They crossed the IDA threshold again in 2007 and stayed above it for the remainder of the sample.

Because they dipped just below the threshold in two years and were already above the threshold for nearly a decade, I chose the first crossing, 1993.

- **South Sudan (excluded):** South Sudan became a country in 2011. It crossed the threshold in 2014 but fell below it in 2015, and there are no data for the rest of the sample, so I did not include it.
- **Sudan (2007):** Sudan crossed the IDA threshold for one year in 1991 and then sustainably crossed in 2007. It fell back below the threshold starting in 2019, but I kept it in the sample because it had stayed above the threshold for over a decade by that point.
- **Syria (excluded**): I had no GNI data for 1987 through 1999 and then Syria was well above the threshold in 2000. They fell below the threshold in 2014 due to the civil conflict and did not report data for 2019 through 2021, so I did not include them in my sample.
- **Tajikistan (2013):** Tajikistan crossed the IDA threshold in 2013 but fell below it in 2016 and remained below it through 2021. Because the country crossed the threshold for three years, I chose to include it in the sample.
- **Ukraine (2003):** Ukraine started out above the IDA threshold from 1989 through 1995 before falling below it in 1996. They sustainably crossed the threshold in 2003.
- **Yemen (2012):** Yemen crossed the IDA threshold in 2012 but fell back below it in 2016. Because the country was above the threshold for four years, I chose to include it in the sample.
- **Zambia (2008):** Zambia crossed the threshold in 2008 but fell back in 2020. Because it was above the threshold for more than a decade, I included it in the sample.
- **Zimbabwe (2012):** Zimbabwe started out above the threshold but fell below it in 1992. They sustainably crossed the threshold starting in 2012.

While Galiani *et al* 2017) employed a sample of 35 countries, my sample expanded this number by more than half to 54. These countries are displayed in Table 1 above. The rapid expansion of the sample is in line with a broader trend of low-income countries experiencing economic growth in recent decades.

Appendix 2. MCC Scorecard Dummy Variables

This appendix explains the method by which I obtained and determined the four MCC scorecard variables: (1) passing half of the MCC scorecard indicators; (2) passing the control of corruption indicators; (3) passing one of the two democratic rights indicators (i.e., political rights or civil liberties); and (4) passing the MCC scorecard (i.e., passing all three of the preceding pass-fail dummy variables). For these dummy variables, I manually extracted this information from the MCC website.[39]

To determine whether a country was passing or failing these four indicators, I examined its scorecard for the three years in which ODA/GNI entered in my regression equations. For example, if a country crossed the IDA threshold in 2010, I extracted the pass-fail dummy for 2011 through 2013. This aligns with the specifications that only account for three years after the country passes the IDA threshold. These data are pulled from the scorecards that are two fiscal years later, i.e., 2011 data are extracted from Fiscal Year (FY) 2013 scorecards. This is because the FY13 scorecards were actually published in late 2012 using 2011 data. This method enables me to determine the pass-fail dummy variables for more than half of the sample (31/52) with three years of scorecard data. However, at the time of retrieval, MCC only made publicly available their scorecards from FY2008 forward, so the official pass-fail determinations on the scorecards only extend back to 2006 for the data on the scorecard.

Given the incomplete historical data, I was forced to make decisions that were less clear-cut for the countries that crossed the IDA threshold towards the beginning of the sample. For eight countries that crossed the threshold between 2003 and 2004, there were one or two MCC scorecards. These were compared to the 2005 data (from FY2007) to confirm their pass-fail status. For 13 remaining countries, I was able to start by looking at the FY08 scorecard, which made pass-fail determinations based on 2006

[39] https://www.mcc.gov/who-we-select/scorecards

data but also included historical data back to 2002. While the medians determining passing and failing are different each year (and not shown on the scorecard), this can give a sense of the trajectory of the previous four years for six of the countries.

For all of the remaining 13 countries, I examined the raw scores for each indicator for the three relevant years compared to the median score for the entire sample. This roughly proxies the MCC pass-fail approach (though for a different set of countries). When combined with the FY08 scorecard (with data back to 2002), this made many of the pass-fail determinations obvious. For example, Equatorial Guinea crossed the IDA threshold in 2001, and it is obvious from both their raw scores relative to the median as well as the historical data on their FY2008 scorecard that they would have failed all four variables. On the other hand, Samoa, Solomon Islands, and Kiribati (the three countries passing the IDA threshold the earliest), all easily pass all four of the variables. Of those that were less obvious, only a handful were difficult to determine. Those decisions are explained here:

- **Indonesia democracy, 1996-1998 (fail):** While Indonesia passes democracy on its FY2008 scorecard and also appears to score well from 2002 to 2005, it does much worse historical in the Freedom House data. It is below the median for 1996 through 1998, and this coincides with the final years of the non-democratic Suharto regime.
- **Honduras half of all indicators, 2000-2002 (pass):** Honduras passes half of the scorecard indicators in FY2008 for the 2006 data. However, it barely is below the median for the MCC-WGI index variable (of four different WGI indicators on the MCC scorecard) for the years 200-2002, but it appears to easily pass more than half of the indicators for 2002 (with ten or more) on the FY2008 scorecard.
- **China corruption, 2001-2003 (pass)**: China does not pass the FY2008 or FY2009 scorecards for corruption, but it does pass in FY2010. It also appears to be barely passing in 2002 on the FY2008

scorecard. Looking further back, it is well above the median for its corruption score in 2001 through 2003 before getting worse in 2004 through 2011 and then improving again.

I further explore whether my results change if some of the less-clear decisions are omitted from the sample in Appendix 5.

Appendix 3. Full Results for First Stage with Interaction Terms and Reduced Form Equation

The full results of the first stage of the 2SLS regression with interaction terms are presented in Table A1. Columns (1) through (8) in the top panel have ODA/GNI as the dependent variable, while columns (9) through (16) have aid interacted with a governance term as the dependent variable.
Importantly, all of the F-statistics exceed the Stock-Yogo critical value of 7.03. This means that the instrument remains relevant when adding the interaction terms.

In the top panel, it is notable that the coefficients on the interaction terms for the governance dummy variables all have a negative coefficient, though none of the estimates are statistically significant. This suggests that better-governed countries are experiencing a larger decrease in aid once they sustainably cross the IDA threshold. This is intuitive as it is likely the same reason that OLS estimates are downwards biased. Aid allocation across countries is endogenous, and donors tend to favor countries where the problems are larger. As noted earlier, this means giving more aid to conflict-affected or unstable countries. Even though a given fragile state may be experiencing economic growth and graduating from IDA eligibility, they may still have other development challenges that may be greater than better-governed countries. As a result, it is plausible that donors are drawing down aid more quickly in better-governed countries because they are not experiencing these other problems.

Table A1. First Stages Estimates with Interaction Terms

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable:	ODA/GNI							
IDA Crossing (Three Years)	-2.1344	-1.8232	-2.9648*	-2.7752**				
	(1.7515)	(1.1373)	(1.5158)	(1.2094)				
IDA Crossing (All Years)					-2.8389	-1.7804	-3.0247*	-3.0963***
					(1.8624)	(1.1973)	(1.5847)	(1.1769)
Crossing*MCC-Half	-1.8073				-0.8220			
	(1.8224)				(2.2389)			
Crossing*MCC-Corruption		-2.6753				-2.9099		
		(1.6391)				(1.9918)		
Crossing*MCC-Democracy			-0.6186				-0.6156	
			(1.7432)				(2.0528)	
Crossing*MCC-Scorecard				-1.5264				-0.7881
				(1.6528)				(1.9661)
Population (log)	-13.0871	-13.0072	-12.5841	-12.8865	-0.7656	-0.9497	-0.4678	-0.5996
	(8.5183)	(8.4623)	(8.6191)	(8.5488)	(7.4725)	(7.3817)	(7.3988)	(7.4770)
First Stage F-Statistic	8.09	8.65	8.29	9.79	13.63	12.20	13.98	14.09
Stock-Yogo Critical Value	7.03	7.03	7.03	7.03	7.03	7.03	7.03	7.03
Observations	1,020	1,020	1,020	1,020	1,598	1,598	1,598	1,598

	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Dependent variable:	Aid*Half	Aid*Corrupt	Aid*Democracy	Aid*Scorecard	Aid*Half	Aid*Corrupt	Aid*Democracy	Aid*Scorecard
IDA Crossing (Three Years)	1.9638***	0.9865*	1.1575***	0.9208**				
	(0.5038)	(0.5167)	(0.4241)	(0.3750)				
IDA Crossing (All Years)					1.8510***	1.5644***	1.3483**	0.8199
					(0.6642)	(0.5861)	(0.5803)	(0.5156)
ODA/GNI*MCC-Half	-5.8483***				-6.3482***			
	(0.9645)				(1.0850)			
ODA/GNI*MCC-Corruption		-6.0308***				-7.3319***		
		(1.2638)				(1.5549)		
ODA/GNI*MCC-Democracy			-5.1693***				-6.1058***	
			(1.0728)				(1.1979)	
ODA/GNI*MCC-Scorecard				-5.6775***				-6.5183***
				(1.2716)				(1.4773)
Population (log)	-6.4468	-11.0387	-5.6279	-4.6743	-0.2998	-3.1971	2.4970	0.2524
	(4.3098)	(8.5703)	(4.3832)	(4.1701)	(3.4220)	(7.0117)	(3.4842)	(3.1738)
Country Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
Year Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
First Stage F-Statistic	19.86	11.72	32.90	34.41	11.99	8.51	23.46	38.48
Stock-Yogo Critical Value	7.03	7.03	7.03	7.03	7.03	7.03	7.03	7.03
Observations	1,020	1,020	1,020	1,020	1,598	1,598	1,598	1,598

Note: First of 2SLS regression of instrument on aid with interaction terms. * p < .1, ** p < .05, *** p < .01.
Source: World Development Indicators, World Governance Indicators, Freedom House, MCC Scorecards.

Finally, I examine the reduced form equation for the IV model, which regresses GDP growth on the IDA crossing instrument and the instrument interacted with the MCC scorecard dummy variables. The results are shown in Table A2.

Table A2. Reduced Form Estimates of GDP Growth on IDA and IDA*Policy Interaction Terms

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Instrument:	**IDA Crossing (All Years)**				**IDA Crossing (Three Years)**			
IDA Crossing	-4.0056*	-0.6928	-3.7467**	-2.5509**	-3.9748*	-0.6045	-3.3634*	-2.2389*
	(2.0619)	(0.8399)	(1.6331)	(1.1139)	(2.0646)	(1.2456)	(1.6871)	(1.2493)
IDA*MCC-Half	3.0650				3.1303			
	(2.1846)				(2.1667)			
IDA*MCC-Corruption		-2.3992				-2.2836		
		(1.7412)				(1.5954)		
IDA*MCC-Democracy			3.0115*				2.4893	
			(1.7810)				(1.8033)	
IDA*MCC-Scorecard				1.5675				0.9673
				(1.3110)				(1.6285)
Population	✓	✓	✓	✓	✓	✓	✓	✓
Country Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
Year Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
Observations	1659	1659	1659	1659	1078	1078	1078	1078

Note: Reduced form OLS regression of GDP growth on IDA instrument and interaction terms, population, and fixed effects.
* p < .1, ** p < .05, *** p < .01.
Source: World Development Indicators, MCC.

Appendix 4. Choice of Dependent Variable

This appendix focuses various other dependent variables that I have included in my analysis but excluded from the main text, except for the baseline OLS estimates in Table 2. This tests relationship between aid and other important measures of economic development, and it also serves as a robustness check on my main IV results in Table 7. This exercise is partially motivated by the finding of Jim and Williamson (2019) that the original Burnside and Dollar (2000) results only are replicable when using the PWT data for GDP growth as the outcome variable of interest (and not the WGI measure of GDP growth).

As seen in the summary statistics in Table 2, there are both similarities and differences in the mean and variance of these four measures of growth. However, Table A3 shows that there is a strong correlation between most of the indicators. The most strongly correlated indicator to my preferred outcome of interest in the main body (GDP growth from WDI) is most strongly correlated with GNI growth from WDI at 0.77. GDP growth data from PWT has a similarly strong correlation with GDP growth data from WGI at 0.68, but it is notable GDP growth is closer to GNI growth from the same source than it is to the same measure from PWT.

Table A3. Partial correlations of dependent variables

	GDP Growth (WDI)	**GDP Growth (PWT)**	**GNI Growth (WDI)**	**Consumption Growth (PWT)**
GDP Growth (WDI)	1.00			
GDP Growth (PWT)	0.68	1.00		
GNI Growth (WDI)	0.77	0.39	1.00	
Consumption Growth (PWT)	0.32	0.33	0.47	1.00

Source: World Bank World Development Indicators and Penn World Tables.

The rest of the correlations in Table A3 between the WDI and PWT are relatively weaker. The weakest correlation is between GDP growth from WDI and consumption growth, but the correlation between GDP growth from PWT and consumption growth is nearly identical. This suggests that the weaker relationship has more to do with the underlying relationship between the two concepts of measurement than it has to do with major differences between the WGI and PWT datasets. The correlation between GDP growth from PWT and GNI growth is the next weakest at .39. The correlation between consumption growth and GNI growth is .47, which is higher than the correlation for GDP growth from either source. This all suggests that there is not a huge difference between the two data sources, particularly for the preferred outcome of interest (GDP growth), so the results should not change dramatically when shifting from one measure of GDP growth to another. However, there may be larger differences in results when using GNI growth and especially consumption growth as the dependent variable.

I choose to use GDP growth from WDI as my preferred dependent variable for several reasons. First, Jia and Williamson (2019) found that the WDI data produce more conservative estimates, and Burnside and Dollar's (2000) results did not hold when GDP growth data from PWT were used. This potentially raises the bar to show that there is a significant relationship. Second, the GDP growth data from WDI have the largest sample size with the fewest missing observations. This is critical as Jia and Williamson (2019) were unable to test the Galiani et al (2017) due to a lack of statistical power. Third, as I will show next, GDP growth data from WGI produce estimates with the most precision – this is likely due to the greater number of observation and the greater amount of variation than the other variables. Furthermore, this is a common choice in the related literature – most studies of aid-growth employ GDP growth. These reasons led me to employ WDI in the main text, though I also conduct robustness checks here to test whether a different choice of dependent variable might have altered the results.

Table A4. IV Estimates with Various Dependent Variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable:	**GDP Growth (WDI)**				**GDP Growth (PWT)**			
ODA/GNI	0.9425**	0.3286	0.7945**	0.5606*	1.1784	1.0897	0.9283	0.9559*
	(0.4357)	(0.4508)	(0.3689)	(0.2997)	(0.7208)	(1.0603)	(0.5789)	(0.5461)
Crossing*MCC-Half	-0.9078**				-0.6887			
	(0.4428)				(0.6784)			
Crossing*MCC-Corruption		0.1415				-0.2827		
		(0.4212)				(0.9327)		
Crossing*MCC-Democracy			-0.6677*				-0.1323	
			(0.3638)				(0.5642)	
Crossing*MCC-Scorecard				-0.3910				-0.3184
				(0.3131)				(0.5300)
Population (log)	1.8380	-0.1754	0.6887	-0.0180	3.9142	2.5991	2.8835	2.8720
	(7.5505)	(6.2621)	(6.7373)	(6.5665)	(9.3336)	(9.4918)	(9.3370)	(9.4065)
Country Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
Year Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
Observations	1,020	1,020	1,020	1,020	979	979	979	979

	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Dependent variable:	**GNI Growth (WDI)**				**Consumption Growth (PWT)**			
ODA/GNI	1.2235	0.7972	0.5516	0.7911	1.0942*	1.6249	1.0799*	0.9411*
	(1.2808)	(0.7258)	(0.6729)	(0.5872)	(0.5981)	(1.1168)	(0.5584)	(0.5203)
ODA/GNI*MCC-Half	-0.5687				-0.9343**			
	(1.1242)				(0.4595)			
ODA/GNI*MCC-Corruption		0.1077				-1.2308		
		(0.5288)				(0.9033)		
ODA/GNI*MCC-Democracy			0.5851				-0.9502**	
			(0.5317)				(0.4484)	
ODA/GNI*MCC-Scorecard				0.3275				-0.9995**
				(0.5122)				(0.4455)
Population (log)	-1.1683	-1.5037	2.3775	-0.6344	-3.5402	-5.8951	-4.3908	-4.8070
	(7.3077)	(8.1116)	(8.4604)	(7.9413)	(7.6572)	(8.3958)	(7.7245)	(7.2544)
Country Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
Year Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
Observations	633	633	633	633	977	977	977	977

Note: 2SLS regression of various growth variables on ODA/GNI and interaction terms. * $p < .1$, ** $p < .05$, *** $p < .01$.
Source: World Development Indicators, Penn World Tables, MCC Scorecards.

Table A4 shows the IV estimates when using the other dependent variables. I only look at the instrument with three post-crossing years, so Columns (1) through (4) are the same estimates as columns (5) through (8) in Table 7 as a reference point to the main findings. As expected, the coefficient on ODA/GNI is larger but less statistically significant when using the alternative dependent variables. All of the coefficients are positive and greater than .55.

For the interaction terms, there is great similarity between GNI growth and GDP growth from PWT when compared to the original estimates on GDP growth from WGI. Every single coefficient on the interaction terms in columns (5) through (8) and (13) through (16) are negative, which suggests that better-governed countries and democracies do worse at translating aid into economic growth. The only exception is that the democracy interaction term in the main text is positive, though statistically insignificant at conventional levels. The main difference for these two dependent variables is that the interaction terms for the consumption growth outcome are negative and statistically significant in three of the four columns. This reinforces the finding that well-governed democracies have a weaker relationship between aid and growth.

Of the alternate dependent variables, GNI growth stands out as the most different. Two of the four interaction terms have a different sign than the original estimates for GDP growth from WGI, though neither one is statistically significant at conventional levels. This variable also had the smallest number of observations, so I do not believe there is room for concern. Unless there were interaction terms that were positive *and* statistically significant, the main results are robust to changing the dependent variable. In sum, the estimates presented here are very similar to the main results with a few minor exceptions that are almost all statistically significant. Thus, I can remain confident that my IV estimates of the causal relationship between aid and growth are reasonably reliable across multiple measures.

Appendix 5. Robustness check when limiting sample based on MCC scorecard dummies.

As shown in Appendix 2, there are several countries for which it is not clear whether they should be considered passing or failing for the three-year period that is being considered in my preferred specification. This could be because they pass or fail the scorecard criteria two out of the three years or because they crossed the IDA threshold before MCC was producing scorecards. In addition, the "snapshot" nature of whether a country passed or failed during that three-year window may not accurately characterize a country if it is rapidly improving or declining in terms of policy performance.

To check whether this affects my results, I went back through the scorecard dummies and omitted the countries that (1) did not pass or fail all three years in years that scorecard data exists; (2) appeared to be rapidly improving or declining on the relevant criteria where scorecard data do not exist for all years; or (3) appear to be very close to the median or threshold where scorecard data do not exist for all years. This essentially parses out an "always passing" group from an "always failing" group. This crude set of three groups (including the indeterminate middle group) also reflect MCC's historical experience in which about one third of countries have always passed, one third go back and forth, and one third have always failed. Not surprisingly, when only looking at three years, the indeterminate middle group is slightly less than one third as it only captures those right on the threshold currently or the very few countries that are rapidly changing, which is relatively uncommon.

Below are the countries for each scorecard that I omitted from the sample for each dummy variable:

- **Corruption:** Albania, Bangladesh, Georgia, Laos, Mauritania, and Pakistan.
- **Democracy:** Albania, Cote d'Ivoire, Mauritania, Nigeria, and Ukraine.
- **Passing half:** Albania, Armenia, Bangladesh, Cote d'Ivoire, Georgia, Indonesia, Laos, Mauritania, Myanmar, Tajikistan, Ukraine, and Uzbekistan.

- **Scorecard (all three):** Cote d'Ivoire and Ghana.

For example, Cote d'Ivoire crossed the IDA threshold during a time of rapid reform and improvement on its MCC scorecard, whereas Albania crossed the IDA threshold in 1999, and the historical data on the MCC scorecard are sparse and low-quality. There are many more countries omitted for the passing half criteria because of the latter reason – the scorecard changed a lot in the early years of MCC, and many of these countries passed well before MCC was established.

Table A5 shows the results of running my preferred specification on this limited sample compared to the original specification from columns (5) through (8) in Table 7 that includes the full sample. The magnitude and sign of the coefficients are all very similar as well as their level of statistical significance. The main exception is for the control of corruption dummy that flips from a positive to a negative coefficient for the limited sample, though it remains statistically insignificant at conventional levels. Otherwise, there is a strong causal relationship between aid and growth for countries that do not pass half of the MCC scorecard indicators, do not pass the democracy hard hurdle, and do not pass the MCC scorecard overall, whereas the causal relationship is not statistically different from zero for countries that do not pass half overall and less strong for countries that pass the democracy hurdle. The control of corruption hurdle remains ambiguous as the interaction term is statistically insignificant. This all suggests that "snapshot" nature of the MCC scorecard dummy variables would not change if it were more flexible and allowed for change over time. That is, when the countries that are rapidly changing or on the threshold of passing are omitted from the sample, the results do not change.

Table A5. IV Estimates Comparing Full Sample to More Limited Sample for MCC Scorecard Dummy Variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dummy Variables	**Full Sample**				**Limited Sample**			
ODA/GNI	0.9425**	0.3286	0.7945**	0.5606*	0.8881**	0.5355	0.7830**	0.5633*
	(0.4357)	(0.4508)	(0.3689)	(0.2997)	(0.3953)	(0.5552)	(0.3841)	(0.3049)
Crossing*MCC-Half	-0.9078**				-0.9097**			
	(0.4428)				(0.4594)			
Crossing*MCC-Corruption		0.1415				-0.0634		
		(0.4212)				(0.5246)		
Crossing*MCC-Democracy			-0.6677*				-0.6176	
			(0.3638)				(0.3817)	
Crossing*MCC-Scorecard				-0.3910				-0.3256
				(0.3131)				(0.3164)
Population (log)	1.8380	-0.1754	0.6887	-0.0180	13.0914	5.0561	2.9714	1.1174
	(7.5505)	(6.2621)	(6.7373)	(6.5665)	(10.8565)	(7.6296)	(7.3065)	(7.6575)
Country Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
Year Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
First Stage F-Statistic	10.8688	8.6335	12.2945	11.6290	11.5418	7.8167	12.7878	10.7715
Observations	1,020	1,020	1,020	1,020	765	879	932	977

Note: 2SLS regression of GDP growth on ODA/GNI and interaction terms for the MCC scorecard. * $p < .1$, ** $p < .05$, *** $p < .01$.
Source: World Development Indicators, MCC Scorecards.

Chapter 4. Micro Evidence on Need and Effectiveness

Do country-level factors make a difference in whether a foreign aid project achieves its intended outcomes? If so, what country characteristics are most important in achieving project outcomes? With over $200 billion spent in Official Development Assistance (ODA) each year, even small improvements in aid effectiveness can have a large impact on global poverty reduction (Collier and Dollar 2002).

This chapter examines the micro evidence regarding if and how the criteria considered in the allocation of foreign assistance translate into project outcomes. These are important questions because donors strive to achieve the greatest development impact possible by spending their scarce resources on the highest-return projects possible. If the evidence does not support the criteria related to need or effectiveness (as proxied by average income, governance, and democracy), then they should revisit those criteria or consider dropping them altogether. That is, if I find that it is more difficult to achieve programmatic results in a particular type of country, then a donor should seriously re-consider whether those countries represent a good investment of resources. On the other hand, if I find that a particular type of country typically achieves greater programmatic success, then donors should consider whether to re-allocate resources to those higher return opportunities.

To answer these questions, I exploit a novel database of project ratings for more than 20,000 projects that are standardized across twelve different donors. In essence, this takes the big question of "does aid work?" out of the equation, because I am using a direct measure of whether a project achieved its objectives. This enables me to look closely at whether and what country factors are influencing project outcomes. My empirical strategy employs a fixed effects model. I first look at the amount of variation in the project outcomes explained by country fixed effects to investigate how much of the difference in project outcomes is determined by macro-level country factors. This provides evidence concerning the relative importance of country selectivity vis-à-vis the other drivers of project outcomes, such as project management. I then utilize a fixed effects model to test the importance of average income, governance, and democratic rights in influencing project outcome ratings. I utilize donor, sector, and year fixed effects to account for differences in donor grading curves, sectoral differences in ratings (to account for inherent difficulties in achieving results in particular sectors that may be more prevalent in one type of country), and economic shocks and other global phenomenon in a given year, among other reasons.

First, I examine whether country-level factors matter for project outcomes. Previous studies showed that country-level factors are of some importance when considering how to allocate a global portfolio of assistance across countries, although not the dominant determinant of project success (Denizer *et al* 2013). I test this by regressing project ratings on country fixed effects and examining the R^2 to determine how much of the variation in project outcomes is determined by country-level factors. Second, I examine whether project ratings are higher or lower, on average, in poor, well-governed, and democratic countries. I examine the factors of (1) country need as measured by average income (GNI per capita), (2) governance as measured by the World Bank's Country Policy and Institutional Assessment (CPIA) and World Governance Indicators (WGI) broadly and control of corruption specifically; (3) democracy as measured by the Freedom in the World index by Freedom House; and (4) the aid allocation models of the World Bank and MCC employing combinations of (1) through (3). I chose the World Bank and MCC because they are the two donors that most prominently use data-driven models for aid allocation and country selectivity, respectively.[40]

[40] For their IDA funding allocation model, the World Bank utilizes measures of average income and CPIA scores. MCC filters out countries based on average income by only working with LICs and LMICs. MCC then creates an annual country scorecard that measures governance and democratic rights. To become eligible for MCC assistance, a country must be poor, relatively well-governed, and democratic.

I arrive at four main findings in this chapter. First, I find evidence for the importance of country-level factors as drivers of project success. I find that country-level factors account for up to one third of the variation in project outcomes. Second, I find that richer countries achieve better project outcome ratings, on average. Third, I find that good governance is a critical driver of project success, though control of corruption is not particularly important. Fourth, I find that democracy is not a key driver of project performance at the micro level, and *non*-democratic countries perform better in terms of project outcomes. In short, good governance is the most important factor, followed by higher average income, corruption is insignificant, and democracy may be a detriment to achieving project outcomes. These findings are mostly consistent across aid sectors with a few minor exceptions, such as that humanitarian projects achieve higher success in poorer countries, on average.

This chapter makes two primary contributions to the economic literature. First, I find that country-level factors account for up to one third of the variation in project ratings, which suggests that country selectivity may be more important than previously understood. Second, I provide micro-level evidence at the project level for the relative effectiveness of well-accepted aid allocation criteria. First, I find that country-level aid allocation criteria do influence project-level outcomes. Earlier studies (e.g., Dollar and Levin 2005) assume that country-level characteristics are critical determinants of aid effectiveness, whereas Denizer *et al* (2013) investigate the macro and micro correlates of project outcomes. The authors found that country-level institutions and macro conditions account for about 20 percent of variation,[41] but I find that this may be an under-estimate. I find that the amount of variation in project outcome rating explained by country-level factors may be one third or more for all donors, and this implies that country selectivity may be more important than previously understood. This finding

[41] Denizer et al (2013) conclude that approximately 80 percent of the variation occurs within countries due to project size, length, management quality, and other *within*-country factors.

emphasizes the importance of considering country-level factors in the allocation of assistance. This is an important finding because if country-level factors did not matter, then it might not make sense to allocate resources selectively, and donors could avoid trade-offs related to need and effectiveness.

Second, I contribute to the limited micro-level literature on aid allocation by examining the relationship between country-level allocation criteria and project-level outcomes. A very limited volume of research has branched out from the aid-growth debates of the early 2000s to comb the micro data for complementary evidence of the impact of good governance and institutions on aid effectiveness. Notably, Dollar and Levin (2005) examine the drivers of success for World Bank projects but are most interested in unpacking competing narratives regarding institutional quality versus geographic factors (e.g., tropical climate) and other factors, such as absorptive capacity. The authors found that the quality of institutions and policy performance are the most critical determinant of project success. They conclude that country selectivity is therefore paramount to ensure that aid resources are having the greatest impact possible – that is, aid should be allocated to poor countries with effective institutions. This chapter makes a related contribution in testing a broader range of allocation criteria employed by donors – that is, whether richer, better-governed, and/or more democratic countries typically achieve greater project outcomes. I find that governance matters, democracy does not, and I confirm there is a trade-off with needs-based criteria as project ratings increase with average income. These are extremely useful findings for donors looking to make their allocation models more evidence based.

My findings have direct implications for aid allocation policy. Foremost, country selectivity matters. I find that country-level factors matter for achieving project outcomes. Although not supportive of the traditional needs-based allocation criteria, it is intuitive that project ratings improve with average income – operating environments in poor countries are more challenging, and the governments have less capacity to work with donors. This does not necessarily mean that donors should allocate concessional financing to richer countries, however. Rather, because the return on investment for projects in richer countries would have to be much greater for them to have the same development impact as projects in poor countries (Kenny 2021), the marginal increase in project ratings achieved by richer countries probably does not meet this threshold. Instead of shifting aid away from the poorest countries, donors might consider taking on more ambitious projects and/or raising the bar on good governance standards for richer countries. I find that governance (broadly defined) is important for project outcomes while democracy is not, and this is supportive of performance-based allocation models. Donors could potentially improve project effectiveness by increasing their focus on governance, while downplaying democracy. This provides support for the World Bank and MCC allocation models, though slight tweaks may still be considered given the somewhat conflicting macro evidence, e.g., MCC's use of their democracy and corruption "hard hurdles" are not supported by my findings.

Data

I use the Project Performance Database (PPD) 2.0 of project outcome ratings as my dependent variable. Honig (2018) compiled 14,000 project outcomes in 178 countries over 50 years to compare the practices of nine different donors. While he examined a different question regarding organizational decentralization, his dataset was a valuable step forward in looking at project outcomes. As part of a subsequent study on donor transparency, Honig *et al* (2022) compiled PPD version 2.0, which extends the PPD dataset multiple years and adds several new donors. The new dataset of 12 diverse donor organizations provides the opportunity to test the effectiveness of various donor allocation criteria across a broad set of donors.

The PPD 2.0 includes data on 21,198 different aid projects from 183 different countries. It is unique among foreign aid databases in that it constructs an overall measure of project success that is a "consistent and comparable measure of performance across projects, sectors, countries, and time" (Honig *et al* 2022). The project success ratings are standardized across donors based on a set of performance measurement guidelines that are established by the OECD, including efficiency, achieving project objectives, development impact, and sustainability.[42] For example, Denizer *et al* (2013) detailed the World Bank's rating process and Bulman *et al* (2016) compare the World Bank's process to the Asian Development Bank's (ADB) rating system, which is "a broadly similar process." Both studies noted that there are built-in checks in both systems whereby independent evaluators at the World Bank and ADB validate every rating.[43] While Bulman *et al* (2016) recognize that there are some differences in the ratings process across institutions, the authors include a full set of donor and sector dummy variables to account for this (as I do below) and ultimately find that there are few significant differences between the institutions in terms of what drives their project outcomes.

The project success rating from PPD 2.0 is employed as the dependent variable in my analysis. Each observation specifies a project rating, donor, recipient country, project sector, and a start, completion, and evaluation year, among other indicators. I merged the data from PPD 2.0 with the independent variables at the country-year level and averaged the explanatory variables over the duration of the

[42] While there are no obvious cross-donor trends (Appendix 1) and only a slight increase in ratings over time (Figure 3), the use of donor and time fixed effects in my empirical strategy will account for time-invariant differences in donor rating systems and any grade inflation over time not related to actual project success.
[43] In addition, one quarter of World Bank projects are subjected to an additional, more rigorous *ex post* evaluation and up to half of ADB projects receive an additional detailed evaluation. Both studies include a dummy for the independently reviewed ratings and find they are no different than the rest of the project ratings.

project.[44] The independent variables come from the World Development Indicators (WDI) and World Governance Indicators (WGI) databases from the World Bank and the Freedom in the World database from Freedom House. The explanatory variable for need is Gross National Income (GNI) per capita. For measures of merit, I utilize the Country Policy and Institutional Assessment (CPIA) from the WDI and a sub-set of the World Governance Indicators (WGI). I use the CPIA's overall score, which is an equally weighted average of the four clusters. From the WGI, I use three indicators that are included in MCC's scorecard to create an index of MCC-relevant WGI indicators. Though the control of corruption indicator also comes from WGI, it is given extra weight and considered separately as a "hard hurdle" on the MCC scorecard, so I also examine it separately here.

In addition to the measures of governance, I also employ data from Freedom House as a measure of democracy. To do this, I create an equally weighted index of the political rights and civil liberties scores for each country. This roughly mimics the MCC scorecard's democracy hard hurdle for which a country must be above the median for their income group for one of these two indicators to be selected as eligible to develop an MCC compact investment. This index generally tests the importance of democratic rights, which have seen conflicting results in previous studies, e.g., positive and statistically significant in Denizer *et al* (2013) but negative and marginally significant in Bulman *et al* (2016) for the Asian Development Bank.

[44] While I did not include three projects with no dates, I did assign start or completion years when they were missing. I did this by assigning the average duration across all projects of six years from start to completion and two years from completion to evaluation. For example (in the most common scenario), if a country did not have a completion year but had an evaluation year of 2000, I assigned it a completion year of 1998. Alternatively (in the second most common scenario), if a project had a completion year of 1998 but no start year, I assigned it a 1992 start year. This was done for less than 5 percent of the sample and should not have huge implications as my measures of average income and governance typically change slowly over time, and the average over the project's duration should not be significantly affected by one or two years of data errantly included or excluded.

Finally, I limit my sample only to developing countries, i.e., low income, lower middle income, and upper middle-income countries (UMICs). I do this by excluding countries that are high income countries (HICs) for the entire duration of the sample. However, this means that there are some countries remaining in the sample that started out as a developing country but have graduated to become a HIC. Thus, I further limit the sample by excluding any remaining country-year observations in which a country's GNI per capita (in constant dollars) is above the UMIC/HIC threshold in FY 2023 (for 2021) of $13,205. This effectively limits the sample to only developing countries and some projects that were implemented in current HICs while they were still developing countries.

Summary statistics

Table 1 presents the summary statistics for my data. I have obtained 14,515 distinct projects in developing countries from the PPD v2.0. These projects have an average rating 4.24 on a scale of one (worst) to six (best). The standard deviation for these scores is 1.11. The data on average income is more limited with only 11,171 observations. The mean for GNI per capita is roughly $2,657 with a standard deviation of $2,519. The minimum observation was about $211, and the maximum was $13,167 (because the sample was cut off at the World Bank's FY 2023 analytical threshold for HICs of $13,205).

Table 1. Summary statistics

	Obs.	**Mean**	**Std Dev**	**Min**	**Max**	**Source**
Project rating	14,515	4.24	1.11	1.00	6.00	PPD
GNI per capita	11,171	2,657.41	2,519.37	211.51	13,166.72	WDI
CPIA overall	6,137	3.41	0.43	1.53	4.44	WDI
WGI Index	13,933	-0.48	0.51	-2.42	1.33	WGI
Corruption	13,911	-0.59	0.50	-1.78	1.54	WGI
Democracy	14,352	0.00	0.98	-1.90	2.00	FH

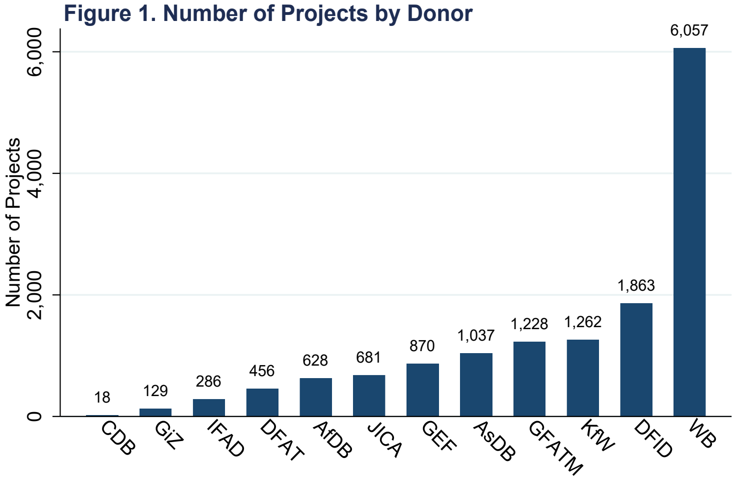
Source: Project Performance Database v2.0; World Development Indicators; Worldwide Governance Indicators; Freedom House.

CPIA scores are also on a one (weak) to six (strong) scale. I have 6,137 observations for CPIA, as the public scores are only available starting in 2005. The average score is 3.41, and the standard deviations is .43. The minimum score is 1.53, and the maximum score is 4.44. I have 13,933 observations for my WGI index. The index has a mean rating of -.48 as the estimates for the three indicators are normalized to zero but for all countries, and my sample only uses developing countries (that tend to have worse governance). The standard deviation is .51, and the range is from -2.42 to 1.33. Corruption follows a similar pattern as it is from the data source using the same methods to calculate it. It has 13,911 observations with a mean of -.59, and a standard deviation of .50. The minimum is -1.78 and the maximum is 1.54. Finally, the democracy index has 14,352 observations with a mean of zero and a standard deviation of roughly one as I normalized the data to zero for my sample. The minimum is -1.90, and the maximum is 2.00.

Descriptive Statistics

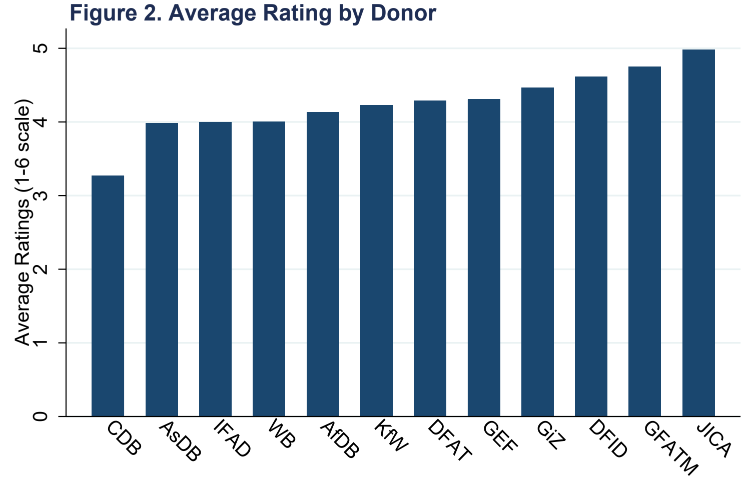
Figure 1 shows the total number of projects in the PPD 2.0 database by donor. The World Bank has the most projects with over 6,000. The United Kingdom's former Department for International Development (DFID, now the Foreign, Commonwealth & Development Office), Germany's national development bank, KfW Group, and the Global Fund to Fight AIDs, Tuberculosis, and Malaria (GFATM) are the next largest in terms of total number of projects with 1,863, 1,478, and 1,326, respectively. The smallest are the Canadian Development Bank (CDB), Germany's development agency, Deutsche Gesellschaft für Internationale Zusammenarbei (GiZ), and International Fund for Agricultural Development (IFAD) with 18, 129, and 290 projects, respectively. Figure 2 shows the average rating of each donor. The weighted average rating across all donors is 4.24. There is some heterogeneity in average project ratings across donors (Figure 2). Most donors fall roughly between 4.0 and 5.0. The Canadian Development Bank (CDB) is the only donor below 4.0 at 3.3, though it has a very small number of projects. The donor with the highest average rating is JICA with a mean rating of 5.0.

Figure 1. Number of Projects by Donor



Note: Total number of projects by donor completed 1995-2027.
Source: Project Performance Database 2.0 (Honig, Lall, and Parks, 2022).

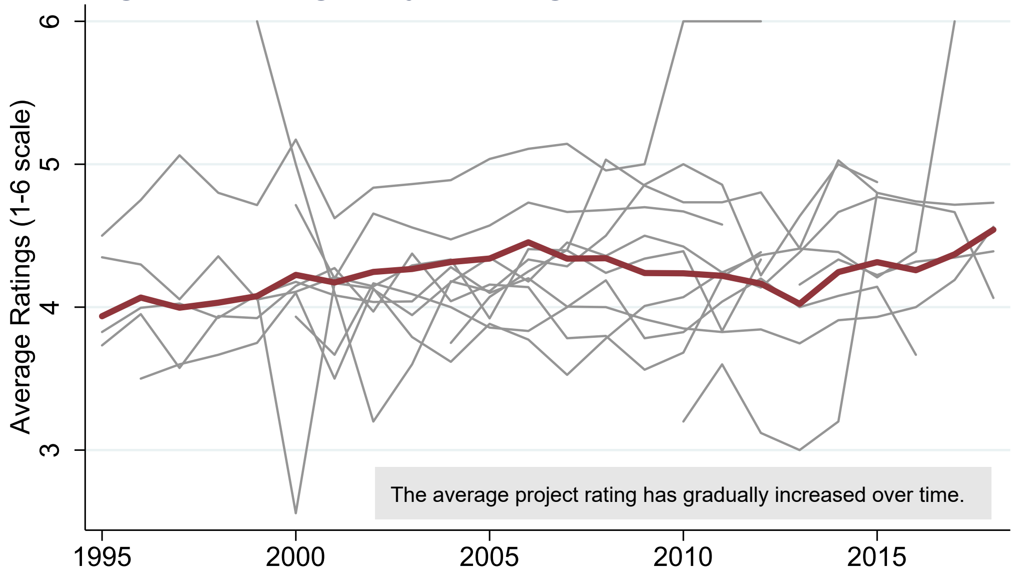
Figure 2. Average Rating by Donor



Note: Average project outcome rating by donor on a scale of 1-6 for projects completed 1995-2027.
Source: Project Performance Database 2.0 (Honig, Lall, and Parks, 2022).

Over time, most donors have stayed relatively stable in their ratings, though there is a modest increase in the average rating over time. Figure 3 shows all ratings from all donors averaged for the year in which the project was completed. Each year has over 350 projects per year from 1995 until 2018 when the number of projects drops to 162. The number of projects continues to drop in the out-years, so those are not included here. After a slight but gradual increase in the average rating from 1995 through 2006, the average rating declined gradually until 2013. In 2014, the average rating again increased, and this was a trend until 2018, which saw the highest average rating of any year shown here. This is partially due to the composition of the donors' reporting projects in the latter years, which is explored in Appendix 1. Most donors have trended in a slightly positive direction in terms of their average rating. This could be due to a general improvement in effectiveness or due to grade inflation. Figure 3 shows all donors on the same graph to give a sense of the dispersion in the trend over time. Appendix 1 examines the trend for each donor's average rating over time in more detail.

Figure 3. Average Project Ratings over Time



Note: Average project outcome ratings (on a scale of 1-6) for each donor by year are show in gray while the weighted average of all projects is show in maroon.
Source: Project Performance Database 2.0 (Honig, Lall, and Parks, 2022).

Overall, it is hard to determine what might be driving these trends, *prima facie*. While some donors appear to have been negatively affected by economic shocks, this is not true across the board, and the all-donor average does not take a significant downturn due to the 2001 recession or the Global Financial Crisis. In general, governance and democracy measures have improved over the medium and long run, so there could be some correlation with improved donor average ratings. However, this may have more to do with smaller sample sizes in recent years.

Finally, Table 2 shows the average rating of project outcomes by categories that I created by grouping together similar three-digit DAC sectors.[45] There is a wide dispersion between the highest rated and lowest rated categories. The category of humanitarian assistance (HA) has the highest ratings (4.79), on

[45] PPD 2.0 lists a three-digit "purpose code" for each project that utilizes the DAC's Creditor Reporting System codes. There are dozens of these sectors in the data and many of them are similar, so I grouped the sectors into broader categories here to distill the sectoral information. For instance, my "Education" category includes Education (110), Education Level Unspecified (111), Basic Education (112), Secondary Education (113), and Post-Secondary Education (114). I use all of the three-digit sectors for my sector fixed effects, however.

average, and this includes the sectors of emergency response (4.93), humanitarian aid (4.75), and food assistance (4.69). This makes sense, because the project objectives are likely related primarily to the distribution of in-kind goods, such as food and shelter. Similarly, global health projects are the second highest rated category (4.33), on average, and project objectives often include easily observable deliverables, such as supplying medicines or providing immunizations. Even though they are often implemented in complex or even crisis environments, humanitarian and health objectives are arguably more straightforward and achievable than the objectives associated with long-term economic and governance objectives that are the result of complex systems largely out of the control of donors.

Table 2. Average Ratings by Category

Category Name	Observations	Mean	Median	Standard Dev.
Humanitarian Assistance	489	4.79	4.80	0.86
Health & Water	2,327	4.33	4.50	1.17
Other	1,963	4.33	4.50	1.12
Infrastructure	2,945	4.31	4.50	1.07
Education	999	4.25	4.50	0.96
Environmental Protection	1,103	4.23	4.00	1.00
Governance & Peace	1,600	4.13	4.00	1.07
Business & Finance	801	4.11	4.00	1.17
Budget Support & Debt Relief	290	4.07	4.00	1.28
Agriculture, Forestry, & Fishing	1,585	4.00	4.00	1.09
Industry, Construction, & Mining	413	4.00	4.00	1.29
Total	14,515	4.24	4.50	1.11

Note: Project outcome ratings by sector category for projects completed 1995-2027.
Source: Project Peformance Database, 2.0 (Honig, Lall, and Parks, 2022)

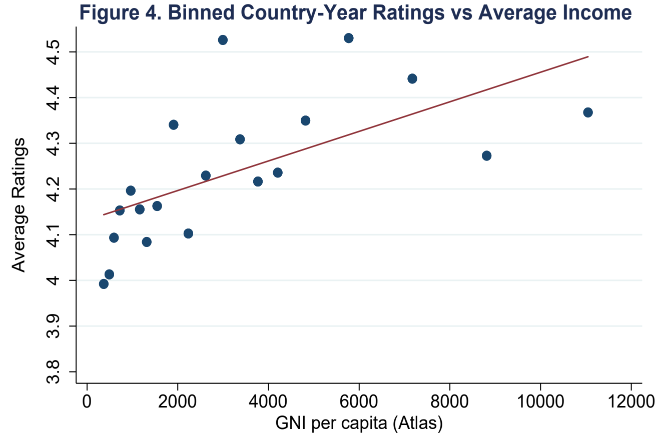
This explanation is also supported by the lowest scoring sectors. Industry, construction, and mining and agriculture, forestry, and fishing are the categories with the lowest averages scores (both 4.00). The next lowest scoring categories (besides budget support and debt relief) are related to business and finance (4.11) and governance and peace (4.13) – both well below the average of all projects (4.24). Tourism (3.71) and government and civil society (3.86) are also among the three-digit sectors with the lowest average rating that fall in other categories. These are all sectors that depend heavily on private sector participation and investment to achieve their objectives or focus on thorny governance issues or policy reforms. Furthermore, both types of projects (private sector and governance) are often complicated by entrenched political economy issues. The heterogeneity of scores across categories and sectors points to the importance of including sector fixed effects in a regression model. This also suggests that it may be worthwhile to examine whether there is uniformity or divergence in the relevance of country-level factors when allocating different types of aid.

Scatter plots

To get a sense of the relationships between project ratings and my variables of interest, I examine binned scatter plots of project outcome ratings and my explanatory variables in Figures 4 through 8. The observations in the scatterplots are an average of each country's project outcome rating across all projects for all donors in a given year, i.e., a country-year rating. The figures then plots the country-year average ratings against each country's average income, governance, or democracy for each country-year. These observations on the scatter plot are then grouped into 20 bins that represent the average of the country-year observations in that bin.

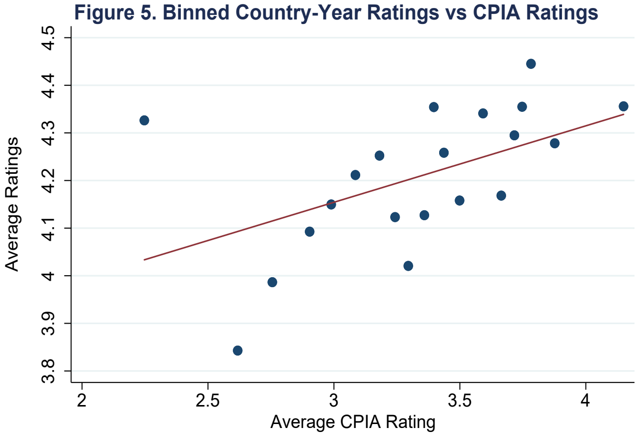
The broad relationships between average project ratings and the explanatory variables are generally what I would expect. Poor countries get much lower average project ratings, governance measures (CPIA, WGI, and corruption) are all positively associated with project ratings, and democracy is slightly positive as well. Notably, there is a very strong relationship between project ratings and governance broadly defined (CPIA and WGI), except for the very worst-governed countries. The better project ratings at the very bottom of governance measures tend to be very fragile states (like Somalia and Eritrea) that are receiving primarily HA. This provides further support for the narrative related to Table 2 that finds that humanitarian projects have been more achievable objectives and thus have higher average rating even in very difficult operating environments.

Figure 4. Binned Country-Year Ratings vs Average Income



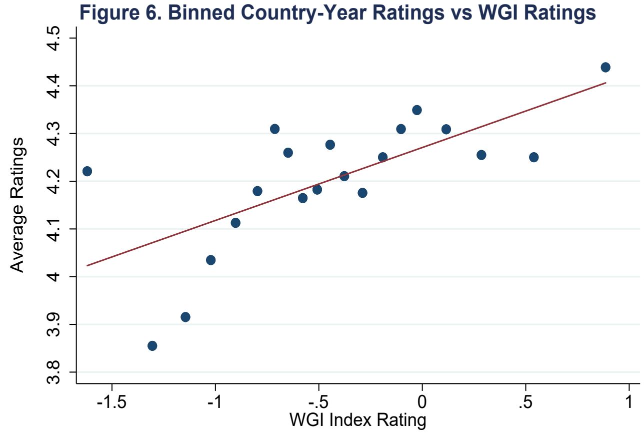
Note: Binned average project outcome rating for each country for each year plotted against that country's GNI per capita in that year.
Source: PPD and WDI.

Figure 5. Binned Country-Year Ratings vs CPIA Ratings



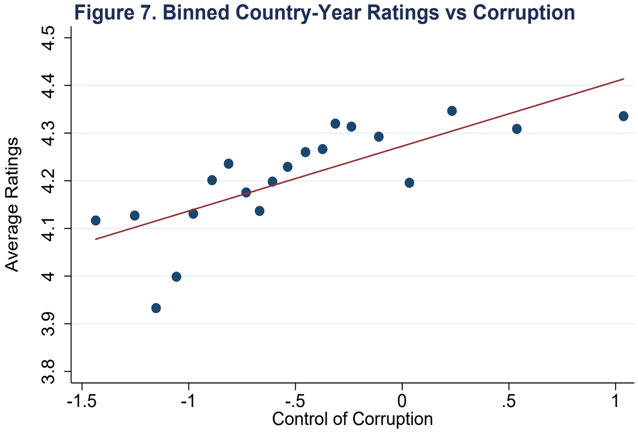
Note: Binned average project outcome rating for each country for each year plotted against that country's CPIA rating in that year.
Source: PPD and WDI.

Figure 6. Binned Country-Year Ratings vs WGI Ratings



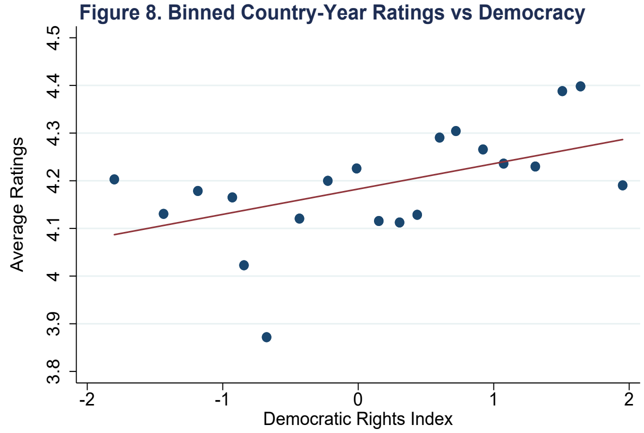
Note: Binned average project outcome rating for each country for each year plotted against the average of country's MCC-relevant WGI indicators in that year.
Source: PPD and WGI.

Figure 7. Binned Country-Year Ratings vs Corruption



Note: Binned average project outcome rating for each country for each year plotted against that country's control of corruption score in that year.
Source: PPD and WGI.

Figure 8. Binned Country-Year Ratings vs Democracy



Note: Binned average project outcome rating for each country for each year plotted against that country's average of its democratic rights indicators in that year.
Source: PPD and WGI.

Looking closer at these relationships, the average project rating is mapped against GNI per capita in Figure 4. In general, the relationship is positive. The poorest countries average about a 4.0 whereas the richest countries average about a 4.5. That is, richer countries achieve better project outcomes, on average. This makes sense. The richer a country, the more resources and capacity it has to work with donors, make programmatic contributions (either in-kind or matching funds), and implement projects alongside projects or with delegated donor funds. However, this may also be interacting with the quality of governance, as better-governed countries also tend to have higher average incomes. I will test this with the full regression specifications proxying the IDA and MCC allocation models.

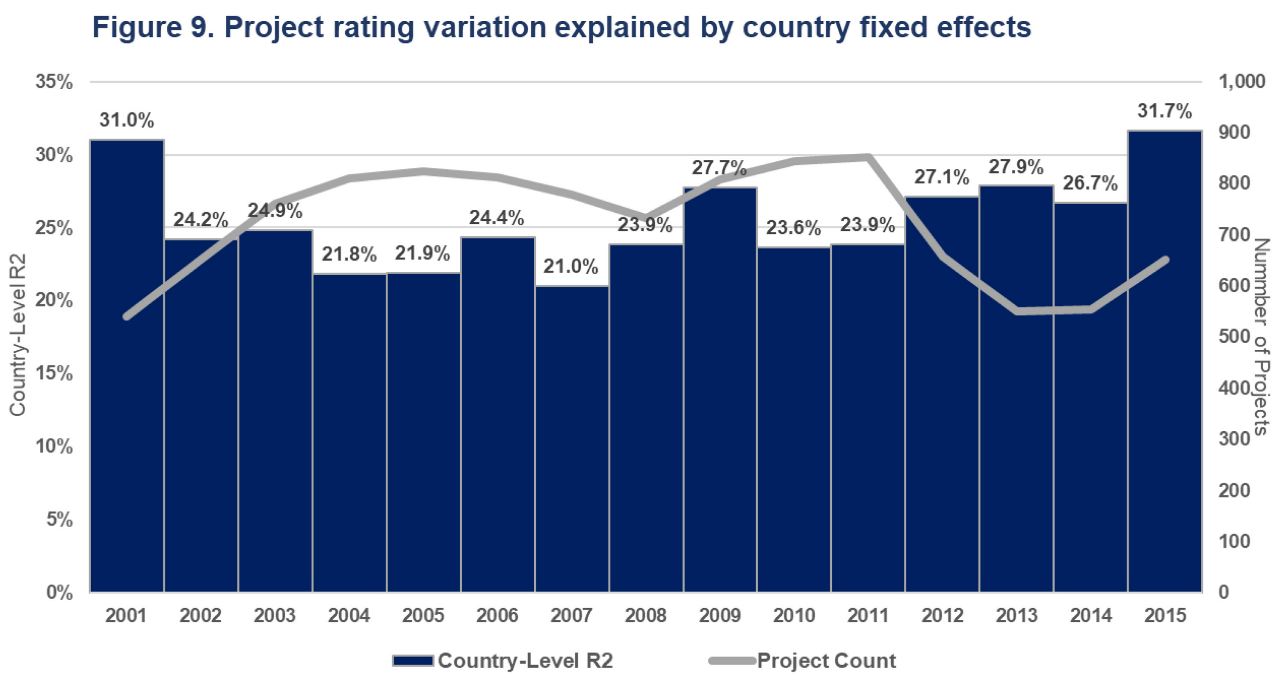
Looking at CPIA scores in Figure 5, I see the relationship that I expected. There is a fairly strong relationship between the quality of governance and project outcomes. The worst governed countries average below a 4.0 whereas the best governed countries are closer to 4.5. Similarly in Figure 6, there is a strong positive relationship between project outcomes and governance as measured by the WGI index. Again, the worst-governed countries average just below a 4.0, whereas the best-governed countries average approximately 4.5. Figure 7 looks specifically at the control of corruption indicator from WGI, and it shows a slightly less pronounced relationship. The best-governed average about a 4.0 while the best-governed countries average about a 4.3 or 4.4. This relationship is a bit weaker than CPIA or WGI. These three figures suggest that governance may be a key factor in determining project outcomes, though the relationship with corruption may be weaker.

Finally, the relationship between democracy and project outcomes is less pronounced but slightly negative in Figure 8. Because the Freedom House scores are one to seven with one being the most democratic, this means that countries that are more democratic have slightly better project outcomes. The lack of a strong relationship is not surprising given the mixed evidence on this criterion.

Country-level variation

One of the novel findings from Denizer *et al* (2013) was that only about 20 percent of the variation in project performance could be explained by cross-country differences. They determined this by regressing country fixed effects (only) on the project outcomes and then examining the R^2. In effect, this parses out the amount of project outcome variation that is explained by *any* aspect of cross-country differences. While I acknowledge the findings that country-level characteristics only account for a fraction of the variation in project outcomes, I instead interpret these findings as confirming the importance of country-level selectivity. Donors are constantly looking to make improvements in the effectiveness of their assistance at the margin, and from that perspective, 20 percent of variation is meaningful. That is, even if only 20 percent of the variation in project outcomes is determined by good governance, this is still an important factor.

In Appendix 2, I first mimic the finding in Denizer *et al* (2011) that country-level variation explains about 20% of the project outcome variation. The R^2s for each year are remarkably similar, even though my data source is not exactly the same as that used in that paper. This gives me confidence that the datasets are reasonably comparable, and the findings are not specific to the original World Bank dataset. I then extend the analysis to my full sample for all donors and count projects differently to test the durability of those results. When including all donors over the full time period counting projects in every year in which they are active, country fixed effects explain only about 12 percent of the variation in project outcomes over the time period (1995-2005) instead of 18-20 percent. This makes sense as there is likely more variation in project outcomes within countries due to different donor approaches. However, because counting active projects multiple times might unnecessarily limit project outcome variation, I also consider projects only in the year in which they are completed.[46] While this changes the interpretation slightly, the R^2 is much larger. The results of combining all donors and only counting a project in the year in which it was completed are shown in Figure 9.



Source: Project Performance Database 2.0.
Note: Figure includes all donors' projects combined. Project rating is only counted included in the year in which it was completed. Figure only includes years with more than 500 projects.

The average amount of project outcome variation explained by country fixed effects in each year is more than twice as large as the sample including all active projects. For the 1995-2005 period, country fixed effects explain about 29 percent of the variation in project outcomes. For the 2006-2020 period, the R^2 increases to over one third. If I only look at the years in which there are more than 500 projects, such as in Figure 11, the average R^2 is 25.4 percent with a range of 21 to 32 percent.

[46] See Appendix 2 for a more in-depth discussion of the trade-offs associated with this method.

Finally, I look at other donors that have been added to PPD 2.0 to see if they are significantly different than the World Bank. Though the sample size is too small to look at individual donors by year, I examine the variation in project outcomes explained by cross-country fixed effects for each donor across all years in the sample. While there were large differences across donors, the average R^2 was 27 percent. Excluding the donors with relatively few projects, the average R^2 was still 15.4 percent for the sample of completion years and 16.9 percent for the active years. Given that fixed effects for the full sample are only capturing unobservable country characteristics, this might be considered a lower bound of the variation in project outcomes that is explained by country-level variation.

Taken together, I both confirm previous findings and find that they may be conservative estimates. My first exercise reinforced the Denizer *et al* (2011) finding that about 20% of World Bank project outcomes are explained by country-level variation. Furthermore, I found that this proportion may be even larger after removing the potential negative bias of counting all active project-years even when including other donors. The amount of variation in project outcome rating explained by country-level factors may increase up to one third or more for all donors using completion years only. This implies that country selectivity may be more important for project outcomes than previously understood, and it emphasizes the importance of considering these factors in the allocation of assistance. In addition, this is an important finding in terms of policy implications for donors because if country-level factors did not matter for project outcomes, then it would not make sense to allocate resources selectively across countries.

Empirical Strategy

I examine whether there are differential project outcomes as a result of country selectivity across the dimensions of need (average income) and effectiveness (good governance and democracy). There are four different dimensions of interest for my research: (1) whether projects are more likely to be successful in richer countries; (2) whether projects are more likely to be successful in better-governed countries; (3) whether projects are more likely to be successful in more democratic countries; and (4) whether the aid allocation models employed by selective donors lead to better project outcomes.

Dollar and Levin (2005) discuss potential sources of endogeneity, including reverse causation (i.e., projects contribute to better institutions) and a "halo effect" whereby the assessments are subjective and countries that are perceived to have effective institutions are also more likely to be perceived to have better projects (regardless of their actual outcome). They dismiss the former on the basis that it is unlikely that successful projects would cause a meaningful and systemic improvement in institutional quality. This is further reinforced by (1) the relatively small size of aid compared to the economy and the fact that I look at levels (in terms of average income) that change very slowly rather than economic growth rates (that have much more variation); and (2) governance measures are "sticky" and multi-faceted, i.e., they don't change quickly and characterize complex systems of governance across many dimensions. Furthermore, from a practical perspective, for the endogeneity concern to be real, the improvements in institutions would have to be vast and immediate for this to show up in the country-level data that is averaged over the life of the project. While the authors do not explain away the halo effect, they conduct both Ordinary Least Squares (OLS) and Instrumental Variables (IV) regressions and produce evidence that the relationship is causal, and the OLS estimates are not substantially biased. I test related concerns that donors are "grading to a curve" within countries in Appendix 5.

Unlike the aid-growth question, the typical endogeneity concerns associated with the allocation decision do not apply here. When looking at the *macro* outcomes of assistance, endogeneity of the allocation decision is extremely problematic since the same rationales for providing assistance will also negatively affect common macro-economic indicators, such as GDP growth. That is, countries are often allocated assistance *because* they are performing poorly economically or they have suffered an economic shock, e.g., a financial crisis. In this case, the amount of assistance and reasons for providing the assistance are neither relevant nor important. These projects were initiated for a variety of reasons across a range of sectors, but the outcome of interest is *only* whether the project succeeded in accomplishing its development objectives – and not a broader measure of economic performance or development outcomes that are influenced by a host of factors, many largely out of the control of donors and even country partners. Because these country factors are plausibly exogenous based on the discussion above, I am able to implement a fixed effects model using OLS regression analysis to determine the drivers of project outcomes.

I first conduct naïve OLS regressions without a full set of controls using estimating equation (1):

$$y_{isdct} = \beta_0 + \beta_1 x_{ct} + \varepsilon_{ct} \qquad (1)$$

$$y_{isdct}: Project\ outcomes\ (i)\ that\ vary\ across\ sector\ (s),\ donor\ (d), country\ (c), and\ time\ (t)$$

$$x_{ct}: Explanatory\ variables\ (need, governance, democracy)\ that\ vary\ across\ country\ and\ time$$

These are essentially partial correlations between project outcomes and each of the allocation criteria and then various combinations of the criteria.

The sample I employ consists of projects for which there is an outcome rating for a project that was completed by 1995. While there are projects extending all the way back to 1956 in the PPD 2.0 database, 1995 is when the World Bank implemented their 1-6 scale, and this seem like a reasonable (though arbitrary) historical cut-off when considering older projects. This includes a handful of projects that are not yet completed (and are estimated to be completed as far out as 2027) yet have an interim evaluation rating. In Appendix 3, I also analyze a more limited sample that only includes projects completed after 2005. I chose 2005 because this is when donors became increasingly focused on foreign aid effectiveness following the Paris Declaration on Aid Effectiveness (OECD 2005). Running the regression for two different periods tests the robustness of my results if the sample is limited to a more current era where donor approaches to project planning and implementation may have shifted.

Then I run an OLS fixed effects model using estimating equation (2):

$$y_{isdct} = \beta_0 + \beta_1 x_{ct} + \Psi_s + \gamma_d + \lambda_t + \varepsilon_{sdct} \tag{2}$$

$$y_{isdct}: Project\ outcomes\ (i)\ that\ vary\ across\ sector\ (s),\ donor\ (d), country\ (c), and\ time\ (t)$$

$$x_{ct}: Explanatory\ variables\ (need, governance, democracy)\ that\ vary\ across\ country\ and\ time$$

$$\Psi_s: Sector\ fixed\ effect$$

$$\gamma_d: Donor\ fixed\ effect$$

$$\lambda_t: Time\ fixed\ effect$$

As before, I continue to utilize the full sample of projects completed in 1995 or later, but also test a more limited sample of 2006-2020 in Appendix 3. As in the partial correlation section, I progress through the various explanatory variables, starting just with average income as a measured of need, then testing governance with both CPIA and WGI separately, then the MCC hard hurdles of control of corruption and democracy separately, and finally proxies for the World Bank's IDA allocation model and MCC's scorecard – average income and CPIA for the World Bank and WGI, corruption, and democracy for MCC.

Finally, I include three different sets of fixed effects. First, I add donor fixed effects, as I showed previously that there is heterogeneity across average ratings across donors (in Figure 2 and Appendix 1), and different donors could potentially grade their projects systematically different – either easier or harder. Second, I add sector fixed effects, as different sectors can be more difficult to produce results than others. For instance, in Table 2 I showed that HA achieved very high ratings, on average, while agriculture projects generally got lower ratings. Third, I add a set of year fixed effects. This will control for any global shocks that might have affected all countries in a given year. This also accounts for the heterogeneity in average scores across the years and the mild grade inflation over time shown in Figure 3. I use the completion year for the time fixed effects, though I find in Appendix 4 that using the mid-point year or the start year does not alter my results.

Results

The results for the partial correlations in equation (1) are presented in Table 3. The relationship between log GNI per capita and project outcomes is positive and statistically significant at the 1% level. At the mean (about $2,657), a 100 percent increase in GNI per capita (approximately one standard deviation) results in a .074 increase in a country's average project rating. This is not a large effect given that a doubling of GNI per capita would take the average income from approximately $2,650 (LMIC) up to about $5,200 per capita (UMIC) – a large change in the level of development of a country that often takes many years or even decades of progress.

The relationship between governance as measured by a country's CPIA score and project outcomes is positive and significant at the 1% level. Given that CPIA is on a scale of 1 (low) to 6 (high), the coefficient can be interpreted as meaning that a one-point increase in a country's CPIA score would correspond to a .18 point increase in project outcomes, on average. While this seems larger than the coefficient on GNI per capita prima facie, a one point increase in the CPIA rating is about 2.5 standard deviations from the mean, which makes the implied increase comparable in size to average income.

The correlation between the WGI index and project outcomes is also positive and statistically significant at the 1% level, though the coefficient is smaller than CPIA. This suggests the positive correlation between governance and project outcomes is robust across two different measures of governance. The control of corruption indicator is also positive (and smaller) and statistically significant at the 1% level. Moving to the democracy measure, I find more of a mixed picture. The correlation between the democracy index and project outcomes is not statistically significant at conventional levels. While the coefficient is negative, it is very close to zero.

Table 3. Partial Correlations of Project Outcome Ratings Regressed on Explanatory Variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Need	**CPIA**	**WGI**	**Corruption**	**Democracy**	**IDA**	**MCC**
GNI per capita (ln)	0.0740***					0.0455*	
	(0.0113)					(0.0235)	
CPIA overall score		0.1765***				0.1882***	
		(0.0313)				(0.0370)	
WGI index			0.1320***				0.2352***
			(0.0184)				(0.0351)
Corruption				0.1094***			-0.0011
				(0.0185)			(0.0335)
Democracy					-0.0111		-0.0875***
					(0.0095)		(0.0116)
Observations	11,171	6,137	13,933	13,911	14,352	4,823	13,900

Note: OLS regression of project outcomes on need and merit variables. * p < .1, ** p < .05, *** p < .01.
Source: PPD v2.0, WDI, WGI, and Freedom House.

I also combine GNI per capita and the CPIA score in column (6) as a crude proxy for the World Bank's IDA allocation model. Although this is a simplification, IDA's allocation model takes into account need and governance (but not democracy). When these two explanatory variables are entered into the regression together, the CPIA coefficient is slightly lower than by itself in column (1), and it is minimally statistically significant at the 10% level. The coefficient for CPIA is roughly the same as column (1) at .19 and remains statistically significant at the 1% level.

Finally, I combine the WGI index, the control of corruption indicators, and the democracy index as a crude proxy for the MCC allocation system. As discussed above, MCC only works with relatively well-governed countries that pass the MCC country scorecard. There is significant overlap between the CPIA and the country scorecards, e.g., government effectiveness, and the scorecards also directly utilize the two elements of the Freedom House score combined here, i.e., civil liberties and political rights. Therefore, although the process is much different for MCC and ultimately selections are made by its Board of Directors, these are the primary determinants that go into MCC's resource allocation model.

The MCC specification in column (7) results in a number of changes to the coefficients from when they are considered separately. The coefficient for the WGI index nearly doubles from column (3) to .24 and is statistically significant at the 1% level. On the other hand, the corruption indicator is no longer statistically significant like in column (4), and the coefficient is negative, though very close to zero. Finally, the coefficient on democracy remains negative, but it becomes statistically significant at the 1% level. These initial findings will be explored further in the next results section.

To summarize this specification's results, average incomes, CPIA scores, the WGI index, and control of corruption are positively correlated with the average project outcome rating, while more democratic countries do not exhibit a statistically significant relationship, on average. While the findings for the IDA allocation model combining income and CPIA scores is supportive of the World Bank approach to resource allocation, the same cannot be said of the MCC scorecard. MCC takes a broader approach to incentivizing good governance and democracy, but these partial correlates initially suggest that the democracy filter may make its projects less effective. Similarly, working only with poor countries could make program success more difficult as well. These are naïve estimates without controls, however, and I now turn to more sophisticated specifications.

Fixed effects model

I now run my fixed effects model for my preferred specification that includes donor, sector, and year fixed effects. I test the robustness of these results to other combinations of fixed effects in Appendix 4, I conclude by examining whether these results are robust across different categories of aid.

Table 4 shows the results of my preferred specification for the fixed effects model. Looking first at the proxy for need in column (1), as measured by the natural log of GNI per capita, I find that the coefficient is nearly 60% larger than the partial correlate when fixed effects for donor, sector, and year are included. The coefficient is now .12 and statistically significant at the 1% level. Again, this implies that a doubling of the average income from the mean would results in an increase in the average project outcome rating of about .12, all else equal.

The coefficients on the governance variables increased even more. Column (2) has a smaller sample size because CPIA data are only available starting in 2005. When the fixed effects are included, the coefficient goes from .18 in Table 3 to .32 here. This relationship is statistically significant at the 1% level. Similarly, in column (3), the index of the three WGI variables on the MCC scorecard (not including corruption) now has a coefficient of .26 with the fixed effects. This is up from .13 without the fixed effects, and it remains statistically significant at the 1% level. Finally, the coefficient for corruption also nearly doubles when fixed effects are added. The coefficient was .11 in Table 3, yet this increases to .21 here in Table 4. It also remains statistically significant at the 1% level. Because the standard deviations for the governance variables are .43-.51, a full one-point increase in those independent variables are relatively larger than a doubling of GNI per capita. This again implies that the .21-.32 increase in average project outcome ratings as a result of a one-point increase in those governance variables is roughly comparable to the coefficient for GNI per capita.

Table 4. Fixed Effects Model of Project Outcome Ratings Regressed on Explanatory Variables

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
	Need	**CPIA**	**WGI**	**Corruption**	**Democracy**	**IDA**	**MCC**
GNI per capita (ln)	0.1178***					0.0358	
	(0.0117)					(0.0239)	
CPIA overall score		0.3158***				0.3418***	
		(0.0308)				(0.0371)	
WGI index			0.2564***				0.3765***
			(0.0187)				(0.0353)
Corruption				0.2067***			-0.0036
				(0.0185)			(0.0327)
Democracy					0.0153		-0.0969***
					(0.0093)		(0.0115)
Donor Fixed Effects	✓	✓	✓	✓	✓	✓	✓
Sector Fixed Effects	✓	✓	✓	✓	✓	✓	✓
Year Fixed Effects	✓	✓	✓	✓	✓	✓	✓
Observations	10,952	5,954	13,692	13,670	14,111	4,656	13,659

Note: OLS regression of project outcomes on GNI per capita with fixed effects for projects complete 1995-2027.
* p < .1, ** p < .05, *** p < .01.
Source: PPD v2.0, WDI, WGI, and Freedom House.

The coefficient for democracy in column (5) does not change much from Table 3. Though it switches from a negative sign on the coefficient to a positive sign, the effect remains very small and statistically insignificant at conventional levels. The standard deviation for the democracy index is close to one, but it does not matter for interpretation here given that the coefficient is not distinguishable from zero.

I now examine the World Bank allocation models in columns (6). Again, the sample size for the specification that proxies the IDA allocation model with average income and governance variables (in the form of the overall CPIA score) is much smaller than the other specifications as it includes the CPIA score, and that data only begin in 2005. Here the governance variable nearly doubles (when compared to Table 3) like the other governance variables on their own, but the coefficient for average income gets smaller and is no longer statistically significant. This is notable because it is the only coefficient in Table 4 that does not get larger when fixed effects are added.

Finally, the results of the specification that is a proxy for the MCC scorecard approach are presented in column (7). This includes the explanatory variables of the MCC-relevant WGI index, the WGI's control of corruption indicator (on its down), and the democracy index – these components roughly mimic the MCC scorecard's criteria of passing half the indicators and passing both the control of corruption indicators and one of the two democratic rights indicators (political rights and civil liberties). Interestingly, the coefficient for the WGI index continues to get larger, while corruption and democracy remain similar to the partial correlations in Table 3. The coefficient for the WGI index grows from .24 in Table 3 to .38 when the fixed effects are added. This is nearly 50% larger than the WGI coefficient in column (3), which was nearly twice as large as in Table 3.

Just as in Table 3, the corruption indicator becomes statistically significant at conventional levels when combined with the WGI and democracy indices. This suggests that the effects of governance on project outcomes are soaked up by the other governance indicators in the WGI index (rule of law, government effectiveness, and regulatory quality). Finally, the coefficient for the democracy index is negative and statistically significant at the 1% level. While the coefficient switches signs and is larger in absolute terms than democracy on its own in column (5), it is roughly the same effect seen in Table 3 when combined with WGI and corruption.

Finally, I examine both the IDA allocation model and MCC scorecard allocation factors for the different categories of aid in Table 5.[47] The results for the IDA allocation model in columns (1) through (10) are broadly consistent with my main findings that the relationship between project rating and average income is positive but not statistically significant, and the relationships between project ratings and CPIA scores are positive and statistically significant. There are notable exceptions, however. Health projects tend to have higher average project ratings in richer countries, while HA tends to have lower project ratings in richer countries. This is the expected relationship, however, the average income coefficient for the IDA allocation model was not statistically significant when pooling all categories in column (6) of Table 4, and it is not statistically significant at conventional levels for any of the other categories of aid. Also notable is that the average income coefficient for HA in column (10) is negative and statistically significant at the five percent level, which suggests that HA projects are more successful in poorer countries, on average. Otherwise, most of the category-level results are consistent with Table 4.

The coefficients for the MCC scorecard allocation factors in columns (11) through (20) are consistent with each other and my main findings that governance is important, corruption is statistically insignificant in most cases, and there is weak evidence that democracy is correlated with lower project ratings. All of the coefficients for the WGI index (of broad governance) are positive, and seven of the ten

[47] Many of these specifications should be interpreted with caution, however, given the small sample size given the limited availability of CPIA data.

specifications are statistically significant at the five percent level or greater. None of the coefficients for corruption are statistically significant at conventional levels. The democracy coefficient is again a somewhat mixed picture, though the coefficient is negative for most of the specifications and only statistically significant when it is negative for four of the specifications at the one percent level. Again, industry (17) and environmental protection (18) do not show any statistically significant relationships at conventional levels. However, these are the exceptions, and these findings generally confirm my main findings that governance is critically important, particularly in health, education, governance, infrastructure, and agriculture.

Table 5. Fixed Effects Model of Project Outcome Ratings Regressed on Donor Allocation Models for Sector Categories

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Education	**Health**	**Governance**	**Infra**	**Business**	**Ag**	**Industry**	**Environ.**	**Budget**	**HA**
GNI per capita (ln)	-0.0501	0.1074**	0.0436	0.0118	0.1434	0.0109	0.4450	0.1556	0.4434	-0.2156**
	(0.0821)	(0.0499)	(0.0594)	(0.0573)	(0.1140)	(0.0796)	(0.3958)	(0.1050)	(1.0039)	(0.0980)
CPIA overall score	0.3048**	0.4533***	0.1659*	0.3896***	0.2393	0.3917***	-0.6105	-0.0903	-1.0322	0.1257
	(0.1404)	(0.0706)	(0.0956)	(0.0935)	(0.2185)	(0.1229)	(0.5758)	(0.1756)	(1.5570)	(0.1219)
Donor Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	342	1314	638	742	218	497	62	272	16	199

	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
	Education	**Health**	**Governance**	**Infra**	**Business**	**Ag**	**Industry**	**Environ.**	**Budget**	**HA**
WGI index	0.2761**	0.4485***	0.3557***	0.4299***	0.4254**	0.4233***	0.2328	0.0795	0.6085**	0.0549
	(0.1159)	(0.0766)	(0.1008)	(0.0781)	(0.1797)	(0.1141)	(0.2547)	(0.1230)	(0.3077)	(0.1527)
Corruption	0.1327	0.0070	-0.0643	0.0732	-0.2312	-0.0510	0.0080	0.0510	-0.1940	0.0350
	(0.1048)	(0.0738)	(0.0958)	(0.0713)	(0.1587)	(0.1023)	(0.2319)	(0.1073)	(0.2864)	(0.1658)
Democracy	-0.1238***	-0.1196***	0.0185	-0.1785***	0.0272	-0.1296***	0.0569	-0.0403	-0.0022	-0.0485
	(0.0373)	(0.0269)	(0.0395)	(0.0221)	(0.0579)	(0.0335)	(0.0798)	(0.0359)	(0.1359)	(0.0665)
Donor Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	967	2992	1539	2785	753	1506	378	1090	252	474

Note: OLS regression of project outcomes on various allocation criteria with fixed effects for different project sector categories.
* p < .1, ** p < .05, *** p < .01.
Source: PPD v2.0, WDI, WGI, and Freedom House.

Taken together, these results suggest that in the context of donor resource allocation models broad economic governance and government effectiveness is critically important for project outcomes. On the other hand, control of corruption is not important, and democracy may actually be detrimental to achieving project outcomes.

Interpreting these findings are tricky, but it has long been observed that good governance and democracy do not always go together – there are both well-governed autocracies and poorly-governed democracies. Furthermore, there are components of the democracy indicators that are likely closely correlated with good governance. For example, one sub-category of the political rights indicator of the democracy index is "Functioning of Government." It is possible that when there is a control for governance (and corruption), this specification becomes a cleaner comparison of autocracies versus democracies, and this negative correlation (though small) suggests that autocracies tend to be more successful at directing donor resources and implementing development projects relative to democracies, all else equal.

Policy implications and conclusion

My findings have direct implications for donors like the World Bank, MCC, or any other donor with the opportunity to pursue a more evidence-based approach to country selectivity or resource allocation. Donors strive to achieve the greatest development impact possible with their resources, so if the evidence does not support the allocation filters of need, governance, or democracy, then donors should revisit their criteria or consider dropping some of them altogether. Below I discuss the implications of my findings for the development policy of donors in terms of resource allocation.

Country selectivity

Importantly, I find that country selectivity matters for achieving project outcomes. Indeed, country-level factors are more important for project outcomes than previously understood. This finding emphasizes the importance of understanding the drivers of project outcomes, so that this can inform donors' allocation of resources. In practice, most donors are not selective in their allocation of resources, or they pursue geopolitical goals or other non-development objectives (Alesina and Dollar 2000). Because up to one third of project outcomes may be determined by country-level factors, donors should pay greater attention to the drivers of project outcomes at the country level. At the very least, they should recognize that their choices may lead to worse project outcomes, on average. The impact of selectivity on aid effectiveness have been dismissed in recent years (McKee *et al* 2020),[48] but I find that it is critical for donors to think strategically and deliberately about how and why they allocate resources and what the implications may be for project outcomes. Particularly with an increased focus on fragile states, it is important to consider the potential downsides that working with poorly governed countries may have. This must be carefully weighed against the greater development need and potential development impact that may be inherent in more difficult environments, which is also a key consideration in thinking through the implications of my findings related to GNI per capita.

Development need

Though the findings are intuitive, the policy implications related to development need are not straightforward. While it may be expected that richer countries perform better in working with donors and implementing projects, this is somewhat problematic from a development perspective in that their development need is less. In other words, richer countries are more able to pay for their own development investments, have greater access to international markets to borrow, and have a greater capacity to repay loans. In addition, as I show in the second chapter, there is much greater need in

[48] McKee et al (2020) review the literature on selectivity and come out unfavorably: "…allocations to 'better-governed' countries were prioritized… [but] the links between aid effectiveness and "better-governed" recipients are not well-supported by the cross-country evidence on aid and growth… Project performance also tends to vary much more by project (within the same country) than by country. On balance therefore, we consider a 'good-governance-focus' indicator of aid effectiveness to be unsound..."

poorer countries across a range of dimensions and measures. Thus, while there is a positive and statistically significant effect of GNI per capita on project outcomes, this does not necessarily mean that donors should allocate concessional financing to richer countries. Given my previous findings, this should be kept in perspective with respect to the potential to achieve broader development objectives.

While it may be *easier* to work with richer countries like UMICs, this may not make sense when the potential for development impact is taken into account. Rather, this finding might even support the argument that aid should be allocated towards *poorer* countries. That is, because the return on investment for projects in richer countries would have to be multiples that of projects in poorer countries for them to be equally effective from a development standpoint (Kenny 2021), the marginal increase in project outcome ratings achieved by richer countries probably does not meet this threshold. That is, the marginal utility of income is much higher in very poor countries (like LICs) when compared to relatively richer countries (like UMICs). Kenny (2021) suggests that the return on investment for aid projects in UMICs would need to be two to four times greater than the same investment in LMICs and eight to 16 times the return in the poorest countries, like DRC or Sierra Leone. The effects of average income on project ratings would likely not clear any of these thresholds.

Instead, donors might consider taking on more challenging or ambitious projects in richer countries, though this may already be the case as I control for sector choice in the fixed effects model. Alternatively, donors could raise the bar on their good governance standards for richer countries. This could result in "picking winners" by supporting countries that are already on a rapid positive trajectory instead of simply rewarding countries for being richer, potentially for exogenous reasons, such as natural resource wealth. This approach is already implicit in MCC's scorecard approach in which the pass-fail thresholds for many (though not all) of its scorecard indicators are determined by the median score for that indicator within a country's relevant income grouping. In practice, this means that the threshold is higher for richer countries.[49]

Good governance

I find that governance is critically important to the success of projects. However, the question becomes how to operationalize this given that the macro evidence conflicts with this finding.

The finding that governance matters for project outcomes is both intuitive and supportive of the World Bank's IDA allocation model that heavily weights policy performance, but it conflicts with the evidence on the aid-growth relationship produced in the last chapter. The World Bank explicitly conditions its allocation of assistance on CPIA scores and previous project performance though its country performance ratings, so it may want to consider more heavily weighting the project performance component, as the connection between governance and growth is less clear. This would allow the World Bank to side-step the aid-growth relationship, while elevating the importance of the governance issues that are most closely related to aid effectiveness (and not determined in advance from Washington).

In addition, the bulk of MCC's country scorecards measure governance-related areas of policy performance. MCC explicitly states that it bases its scorecard indicators on the policy areas that matter most for economic growth, but the findings from my last chapter put the role of aid in this relationship into question. However, the finding related to governance in this chapter lends strong support for a performance-based allocation model at the micro level on the basis of good governance broadly defined. This provides cover for MCC to defend its scorecard approach, but it should drop its rhetoric

[49] To the extent that this is already a practice by some donors in the sample (which does not include MCC), it would have been picked up by the donor fixed effects.

related to the aid-growth relationship. In addition, there is a potential incentive effect that is not accounted for here – by allocating assistance to better governed countries, this may induce countries to implement policy reforms to access more concessional financing. This is somewhat speculative, but it is also supported by numerous anecdotes of the so-called "MCC Effect" as well as a study by Parks and Rice (2013) that shows that partner country leaders are aware of the MCC selectivity model and make reforms in response to this incentive. This is yet another argument that MCC can use to continue its scorecard approach if it is disinclined to make major changes to its selectivity model.

However, MCC's use of the control of corruption indicator is not supported by my findings. Because MCC makes passage of the corruption indicator one of its must-pass "hard hurdles," this indicator carries disproportionate weight on the MCC scorecard. Though the coefficient in my preferred specification is positive and statistically significant when corruption is considered on its own (though smaller than the other governance variables), when it is considered alongside a broader measure of governance and democracy, it no longer has a significant effect on project outcomes. If MCC were only concerned with achieving project outcomes, this evidence suggests that they should potentially drop control of corruption from the scorecard. However, anti-corruption is a politically popular topic with Congressional stakeholders, so MCC may be hard pressed to even drop it as a hard hurdle. Nonetheless, MCC should reflect on what it achieves by including corruption as a hard hurdle, particularly since there is evidence that the indicator is an imprecise measure of corruption, and the pass-fail designation is a blunt instrument that unnecessarily punishes countries that are not statistically different from each other in terms of their score (Dunning, Karver, and Kenny 2014).

Democracy

Finally, the focus on democracy as an allocation criterion may be wasted effort when it comes to aid effectiveness. Prioritizing democracies may actually make achieving project outcomes more difficult. There is no case on efficacy grounds to work only with democracies – and perhaps the opposite – so other rationales would have to trump effectiveness concerns. For instance, a donor could argue that there is an incentive effect, though there is even less evidence here than on the governance front. Additionally, a donor could argue that they are taking a values-driven approach and that donors should not support authoritarian regimes that do not extend full rights to their people. This is not based on a (economic) development rationale, and one must also believe that democracy is desirable – this is generally uncontroversial in the West but might be contested elsewhere.

However, a donor could argue that aid allocation on the basis of democracy could help to provide a "democratic dividend" to emerging democracies to solidify their legitimacy. This is both dependent on the ability of the donor and country partner to achieve success with the project and for the project to have an economically meaningful development impact that is visible to a country's populace. Again, this is more speculative and may not be supported by empirical evidence. Ultimately, donors often take a values-driven approach over a data-driven approach to aid allocation, so some form of a democracy filter is not uncommon. Regardless, even if donors consider democracy a desirable trait in their country partners, they should be clear-eyed in considering the potential trade-off for projects outcomes.

Conclusion

I produce four main findings in this chapter. First, I find support for the importance of country-level factors as drivers of project success. This is an important finding for donors because if country-level factors did not matter for project outcomes, then it would not make sense to allocate resources selectively across countries. Second, I find that richer countries achieve better project outcome ratings on average, but this still may not provide a compelling case to redirect resources to richer countries. Third, I find that good governance is a critical driver of project success, though control of corruption is not particularly important. Fourth, I find that democracy is not a key driver of project performance at the micro level, and *non*-democratic countries may actually perform better.

Together, these four findings support an increased focus on country selectivity that should be taken up by other donors beyond the ones cited here. While some donors may already implicitly consider governance, they could potentially improve their project outcomes by downplaying democracy (or corruption in the specific case of MCC). As a bonus, donors may get extra mileage out of an incentive effect if they were more explicit and disciplined in setting criteria that favor good governance over other factors in their allocation decisions, and this would have a positive reinforcing effect on project outcomes as well. The evidence presented here provides support for the World Bank and MCC models, though slight tweaks may still be warranted given the somewhat conflicting macro evidence.

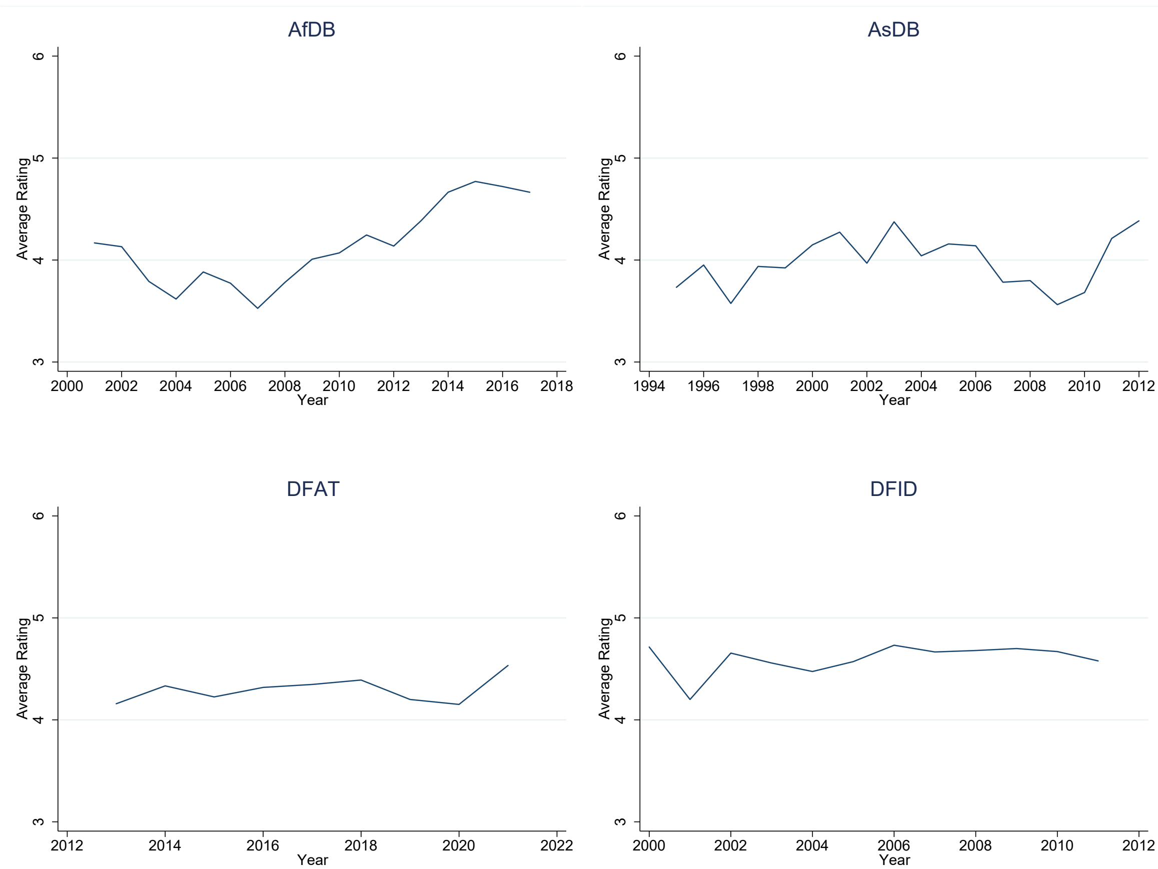
Appendix 1. Trends in average ratings by donor over time

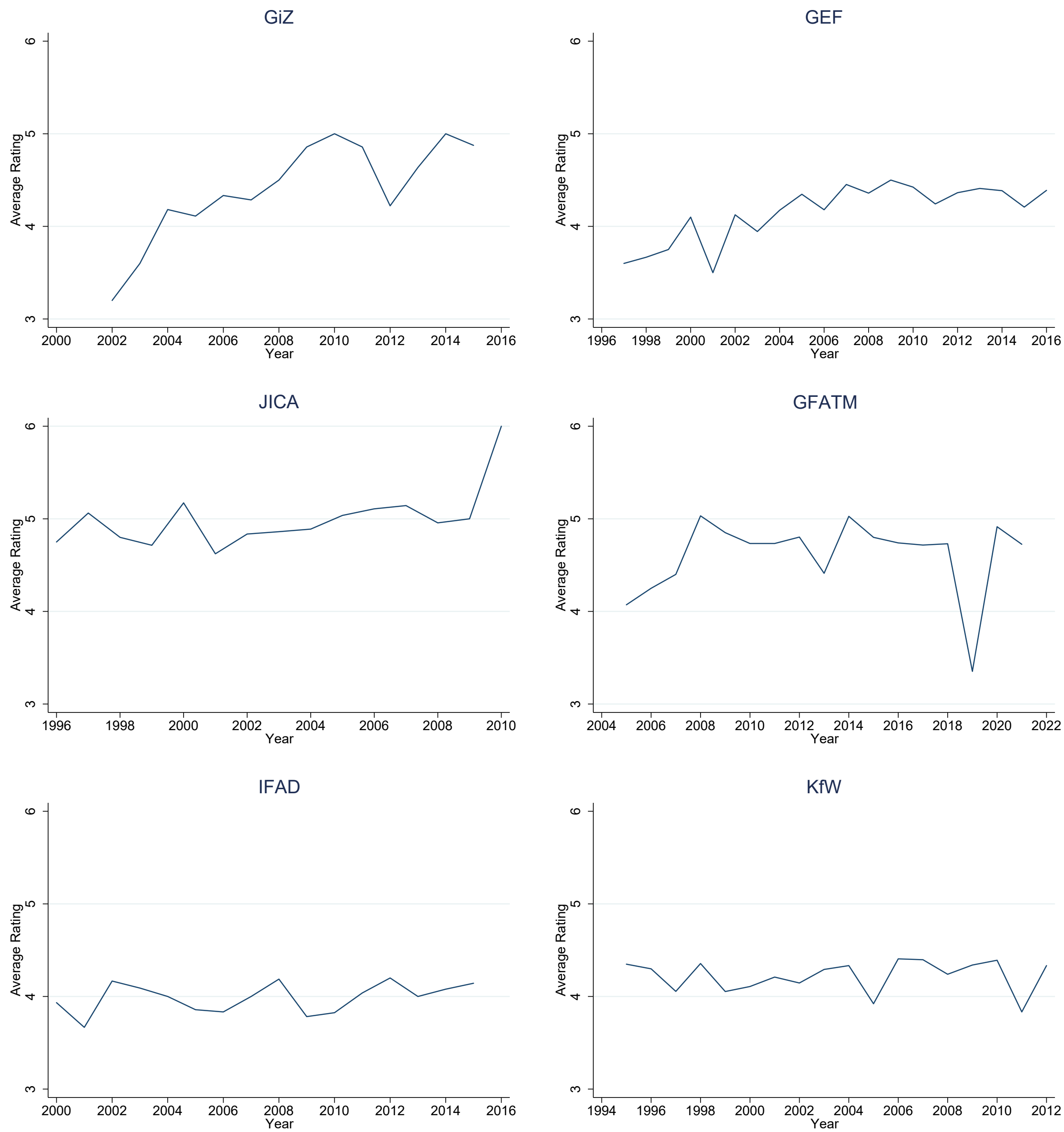
Figure A1 shows the average ratings for each donor across the years in which the donor was most active – each graph removes any years on the tails in which a donor had three or fewer projects.

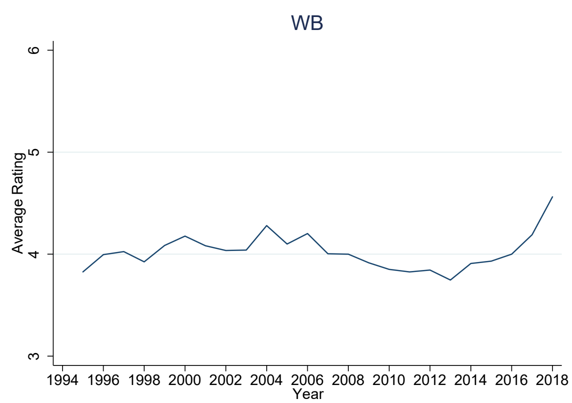
The AfDB averaged 4.13 across all projects for all years. They started around that average in 2001-2002 but dipped below 4.0 between 2003 and 2009, and then stayed above that mark through until 2018. Similarly, the AsDB averaged 3.98 and hovered around 4.0 from the late 1990s through 2006, then dropped off around the Global Financial Crisis in 2007-2010 before increasing again in 2011-2012. CDB only has 18 projects total and only one year (2014) in which they have more than three projects, so I do not display their time trend. DFAT remained remarkably consistent around their average of 4.29 from 2013 through 2021. DFID also performed consistently around their average of 4.6 except for a dip in 2001. GEF starts out below 4.0 in an initial period between 1997 and 2003, but then steadies around their average of 4.3 for the remainder of 2005 through 2016. The GFATM had the most volatile average score, and it appears to be counter-cyclical with increases around the Global Financial Crisis in 2007-2008 and again around the COVID pandemic in 2020-2021. Given the shorter-term nature of health interventions and the fungibility of aid dollars, it is possible that when country governments' fiscal resources were stretched due to other crises, GFATM was able to back-fill higher-return projects that might have been financed by the government otherwise.

GiZ shows the most marked increase over time. They start out well below 4.0 in 2012 before quickly rising to 5.0 by 2010. They remained around 5.0 through 2015 except for a dip in 2012. In contrast, IFAD remains fairly low and at their average over time, around 4.0, for their entire sample from 2000 through 2015. At the other end of the spectrum, JICA remains high throughout their entire time period, hovering around 5.0 until 2010, the last year with sufficient data available, when they shoot up dramatically to about 6.0 – the highest average rating for any donor-year across all years and all donors. KfW is very consistent around their average of 4.2 except for two down-years in 2005 and 2011. If not for this, there would be a very slight upward trend. With the largest sample size, the World Bank has less year-to-year volatility, particularly from 1995 through 2012, when the average rating hovers around 4.0. After a low average in 2013, WB saw a consistent increase from 2014 through 2018 with their highest scores being seen in 2017 and especially 2018.

Figure A1. Average Ratings in Active Years for Each Donor







Note: Average project outcome rating by donor over time on a scale of 1-6 for projects completed in active years where the number of completed projects exceeds three.

Source: Project Performance Database 2.0 (Honig *et al*, 2022).

Appendix 2. County-level variation

In this Appendix, I explore the extent to which country-level variation drives project outcomes. Denizer, *et al* (2013) found that only about 20 percent of the variation in project performance could be explained by cross-country differences, and they determined this by regressing country fixed effects (only) on the project outcomes and then examining the R^2. In effect, this parses out the amount of project outcome variation that is explained by *any* aspect of cross-country differences.

I first mimic Table 6 of the Denizer *et al* (2011) working paper in which they report the results of their exercise. My analysis is not an explicit replication, however, as I am using a slightly different database. While the World Bank project outcome data are from the same source, they have been refined and updated over time for use in the Project Performance Database. I then extend their World Bank-specific analysis to subsequent years and count projects differently (using the year in which they are completed only instead of any year in which they are active) to test the durability of their results. Then I look at other donors that have been added to PPD 2.0 to see if they are significantly different than the World Bank.

Table A1 compares the original World Bank findings to the PPD data starting in 1995, and the R^2 for each year is remarkably similar. [50] The R^2s are all within .03 of each other until 2004 and 2005. In the later years of the World Bank data with smaller sample sizes, the R^2 starts to diverge from PPD. This is a trend that is also seen in the PPD results in the later years of the sample, 2018-2020. While the averages for the entire 1995-2005 period are all within .03 of each other, if I only look at the years with a larger sample size (e.g., above 1,000 observations), the R^2 is almost identical – just .0064 apart from 1995 to

[50] PPD 2.0 has more observations in every year except the first year, 1995. This is likely because projects are not included until they have an evaluation with a rating. Projects that began in the later years did not have an evaluation when the World Bank data were analyzed, though these would get picked up by PPD 2.0 for this period, which was recently released.

2002. This gives me confidence that the datasets are reasonably comparable, and the findings are not specific to the original World Bank dataset.

Table A1. R^2 of Country Fixed Effects Regression

Year	World Bank	Obs.	PPD 2.0	Obs.
1995	0.1820	1,983	0.1869	1,769
1996	0.1720	1,416	0.1669	1,814
1997	0.1790	1,405	0.1664	1,845
1998	0.1780	1,400	0.1867	1,879
1999	0.1700	1,346	0.1984	1,835
2000	0.1590	1,291	0.1863	1,808
2001	0.1700	1,178	0.1869	1,769
2002	0.1910	1,051	0.1734	1,728
2003	0.2190	856	0.1706	1,722
2004	0.2790	641	0.1499	1,728
2005	0.3340	438	0.1502	1,696
1995-2002 Average	0.1751	1,384	0.1815	1,806
1995-2005 Average	0.203	1,182	0.1748	1,781

Note: Amount of variation in the project outcome explained by country fixed effects for active project years.
Source: Denizer, Kaufmann, and Kraay (2011) and Project Performance Database 2.0 (Honig, Lall, and Parks, 2022).

However, the approach to counting active projects multiple times might introduce a downward bias if projects with longer durations receive similar ratings over time, on average. Indeed, Denizer *et al* (2013) find that longer projects do more poorly, so this might reduce the amount of variation overall, which in turn reduces the amount of the variation in the project outcomes that is explained by country-level variation. Instead, since data on completion years are available, I conduct a similar exercise but only count projects in the year in which they are completed. This should remove any potential bias related to the number of times a project is counted.[51]

Table A2. R^2 of Country Fixed Effects Regression by Year

Year	Active Years	Obs.	Completion Year	Obs.
1995	0.1405	3,447	0.3677	377
1996	0.1294	3,681	0.3519	357
1997	0.1219	3,890	0.3274	415
1998	0.132	4,109	0.2745	453
1999	0.1285	4,268	0.2931	474
2000	0.1279	4,443	0.3458	490
2001	0.1252	4,659	0.3101	541
2002	0.1183	4,752	0.242	651
2003	0.1144	4,837	0.2485	761
2004	0.102	4,925	0.2184	810
2005	0.0926	4,920	0.2189	824
2006	0.0935	4,775	0.2435	813
2007	0.0962	4,622	0.21	778
2008	0.0957	4,486	0.2385	733
2009	0.0978	4,331	0.2772	808
2010	0.1071	4,109	0.2362	845
2011	0.1233	3,693	0.2385	853
2012	0.1238	3,103	0.2713	657
2013	0.1453	2,593	0.2785	550
2014	0.1703	2,141	0.267	555
2015	0.1839	1,744	0.3166	652
2016	0.1015	1,173	0.3236	429
2017	0.2964	744	0.3899	362
2018	0.4443	382	0.4529	133
2019	0.5528	249	0.7753	52
2020	0.4906	197	0.4986	156
1995-2005 Average	0.1212	4,357	0.2908	559
2006-2020 Average	0.2082	2,556	0.3345	558
1995-2020 Average	0.1714	3,318	0.3160	559
2001-2015 Average	0.1193	3,979	0.2543	722

Note: Amount of variation in the project outcome explained by country fixed effects for active project years and completion years.
Source: PPD 2.0.

[51] This is a methodological choice for which the right approach is not obvious. The Denizer *et al* (2011) approach counts projects in all active years, which means most projects are counted at least five times. The strength here is that this spreads out its

In Table A2 I present the results of a similar analysis, except that this includes all donors in my sample – not just the World Bank. I also present results for both all active project-years as well as only the project completion year, which is why the number of observations is much lower in the latter. The active years results in a much lower R^2 when all donors are included with only about 12 percent of variation being explained by country-level variation over the time period (1995-2005) instead of 18-20 percent. This makes sense as there is likely more variation in project outcomes within countries due to different donor approaches. Though this complicates this exercise, it is addressed by donor fixed effects in my fixed effects model.

However, the R^2 is much larger when only looking only at a project's completion year. The average amount of variation explained is more than twice as large, and in every single year, the R^2 is larger for the completion year than the active years. For the 1995-2005 period, country fixed effects explain about 29 percent of the variation in project outcomes. For the 2006-2020, the R2 increases to over one third. For the combined period of 1995 through 2020, the R^2 is about 32 percent. Some of this average is driven by very large R^2s in the out-years that are likely driven by the small sample sizes. If I only look at the years in which there are more than 500 projects, however, the average R^2 for those years (2001-2015) is still 25.4 percent with a range of 21 to 32 percent.

rating over the life of the project, while the downside is that this has an inherent bias towards counting longer projects more times – projects that tend to be more difficult and low-scoring (and so this pattern translates into less variation in ratings). My approach only counts projects in the year they are completed. The advantage is that this eliminates double-counting, while the downside is that the project was influenced by country-level factors in previous years, but it does not account for that. Fortunately, all of my explanatory variables are "sticky" and do not change rapidly, so I chose to favor eliminating the double-counting bias.

Finally, I look at individual donors. Though the sample size is too small to look at individual donors by year, Table A3 presents the variation in project outcomes explained by cross-country fixed effects for each donor across all years in the sample. Like in the previous tables, the donors with very small sample sizes have a very large R^2, particularly CDB and GiZ (and IFAD to a less extent). The larger donors are worth examining, however, while keeping in mind a caveat about how pooling the entire sample changes the interpretation of the results slightly.[52]

Table A3. R^2 of Country Fixed Effects Regression for Each Donor

Donor	Active Years	Obs.	Completion Year	Obs.
AfricanDB	0.1989	4,771	0.1960	627
AsianDB	0.1562	7,457	0.1558	1,037
CDB	0.5873	159	0.6436	18
DFAT	0.0862	3,133	0.0649	451
DFID	0.1418	8,683	0.1003	1,866
GEF	0.1619	5,989	0.1594	874
GFATM	0.2623	6,757	0.2428	1,231
GiZ	0.6506	1,062	0.6539	129
IFAD	0.4266	2,377	0.4261	286
JICA	0.2205	4,337	0.1961	684
KfW	0.1866	7,254	0.1690	1,270
WB	0.1058	34,335	0.1056	6,091
Average	0.2654	7,193	0.2595	1,214

Note: Amount of variation in the project outcome explained by country fixed effects for each donor in active and completion years.
Source: PPD 2.0.

There were large differences in the R^2s across donors. For all active years of a project, the average R^2 was .27, but this ranged from .09 (DFAT) all the way up to .65 (GiZ). Setting aside the small sample size donors, the highest R^2 was for GFATM at .2623. Looking only at the completion year of projects paints a

[52] The extended duration of my sample means that country characteristics change over time – from governance and average income to population and natural resource wealth. While country-year fixed effects are a useful snapshot of how much country-level differences affect project outcomes in a given year, applying them across the pooled sample changes the interpretation. That is, the R^2 becomes a measure of how much *unobservable* country characteristics (that don't change over time) affect project outcomes.

very similar picture. Setting aside the donors with few projects, the R^2 ranged from .065 to .243 - again DFAT and GFATM, respectively. At the extremes, this may have something to do with the country presence of these donors. For instance, DFAT has a limited regional presence, largely operating in small island states in the Pacific. Thus, it is not surprising that country-level characteristics are not a huge driver of project outcome variation for those donors. Excluding the donors with relatively few projects (CDB, GiZ, and IFAD), the average R2 was still 15.4 percent for the sample of completion years and 16.9 percent for the active years.

Given the caveat that the country fixed effects for the full sample is only capturing unobservable country characteristics, this might be considered a lower bound of the variation in project outcomes that is explained by country-level variation.

Appendix 3. Testing robustness to different sample time periods

In this appendix, I test a more limited sample that only includes projects completed after 2005 and before 2021. The timing of this more limited sample is driven by the greater donor focus on foreign aid effectiveness that began with the Paris Declaration on Aid Effectiveness in 2005 (OECD 2005). This period also coincides begins in the same year that the CPIA scores began to be published. I also do not include projects completed past 2020, as this is the last year for which a full set of explanatory variables are available. This limited sample period is reported alongside the full sample that is reported in the tables earlier in the main body of the chapter.

This comparison limits the sample to the “aid effectiveness era,” and it also sets up a useful comparison to test the generalizability of the more recent sample over a longer time period. Other studies in this space, such as Burnside and Dollar (2000), have been criticized for being particularly sensitive to the sample selection, e.g., by Jia and Williamson (2019), so this is an important test of robustness.
Table A4 shows the partial correlations across the various specifications without fixed effects. The top panel is the same as Table 4 above to compare against the bottom panel of the more limited sample. The coefficient for GNI per capita is .074 for the full sample, but it is only .036 for the limited sample – both are statistically significant at the 1% level. The coefficient for the limited sample is less than half the size of the full sample. The coefficient for CPIA is slightly smaller for the limited sample and both are statistically significant at the 1% level. The WGI index sees a large decrease as it is about 40% smaller. Again, both remain statistically significant. The coefficient for corruption is also much smaller, going from .109 in the full sample to .051 in the limited sample. Both are statistically significant at the 1%

level. Finally, the coefficient for democracy gets larger in absolute terms, increasing from -.011 to -.047. Only the limited sample is statistically significant, however, at the 1% level.

Table A4. Partial Correlations of Project Outcome Ratings Regressed on Explanatory Variables for Different Time Periods

	Full Sample (1995-2027)						
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
GNI per capita (ln)	0.0740***					0.0455*	
	(0.0113)					(0.0235)	
CPIA overall score		0.1765***				0.1882***	
		(0.0313)				(0.0370)	
WGI index			0.1320***				0.2352***
			(0.0184)				(0.0351)
Corruption				0.1094***			-0.0011
				(0.0185)			(0.0335)
Democracy					-0.0111		-0.0875***
					(0.0095)		(0.0116)
Observations	11,171	6,137	13,933	13,911	14,352	4,823	13,900
	Limited Sample (2006-2020)						
	(8)	**(9)**	**(10)**	**(11)**	**(12)**	**(13)**	**(14)**
GNI per capita (ln)	0.0363***					0.0436*	
	(0.0139)					(0.0242)	
CPIA overall score		0.1669***				0.1832***	
		(0.0321)				(0.0379)	
WGI index			0.0797***				0.2215***
			(0.0227)				(0.0435)
Corruption				0.0507**			-0.0145
				(0.0231)			(0.0429)
Democracy					-0.0473***		-0.1118***
					(0.0117)		(0.0147)
Observations	6934	5624	8263	8268	8283	4480	8263

Note: OLS regression of project outcomes on various allocation criteria with fixed effects over two different periods.
* p < .1, ** p < .05, *** p < .01.
Source: PPD v2.0, WDI, WGI, and Freedom House.

Shifting to the donor allocation models, the coefficients are more similar between the samples than the isolated covariants. The IDA allocation model in columns (6) and (13) are virtually unchanged. The GNI per capita coefficient goes from .046 to .044 and remains statistically significant at the minimal 10% level. The coefficient for the CPIA overall score goes from .189 to .183 and remains statistically significant at the 1% level. These are both with the standard error of each other and therefore statistically indistinguishable Similarly, the MCC scorecard proxy does not change dramatically either. The coefficient on the WGI index goes from .235 to .221 and remains statistically significant at the 1% level. The coefficient on democracy is slightly negative and remains statistically insignificant at conventional levels. The coefficient on democracy changes the most going from -.088 to -.111 and remaining statistically significant at the 1% level.

Although some of the coefficients are significantly smaller in the limited sample, this does not change the broad interpretation of the results. Indeed, not a single coefficient changed signs. If these results hold up in comparing the samples of the specifications with a full set of fixed effects, I can reasonably claim that the results are robust across the two time periods.

The results comparing the samples across two different time periods are displayed in Table A4. Again, the top panel is the same as Table 5 above to compare against the bottom panel of the more limited sample. All of the specifications include the full host of fixed effects for donor, sector, and year.

With the fixed effects included in the specification, the coefficients are all much more similar across the time periods than the partial correlations. The coefficient on GNI per capita decreases from .118 to .102, which is just outside of one standard error of each coefficient. The coefficient on the CPIA score is nearly identical across the two time periods, going from .316 in the full sample to .311 in the limited sample. The coefficient for the WGI index decreases slightly more from .256 to .226, but again this is just outside of the standard error for each estimate. The decrease in the coefficient for corruption is slightly larger, dropping from .207 to .162. All estimates in columns (1) through (4) for the full sample and columns (8) through (11) are statistically significant at the 1% level. The coefficient for democracy remains statistically insignificant at conventional levels, but it switches signs from .015 to -.002.

The estimates for the donor allocation models in columns (6) and (7) for the full sample and (13) and (14) for the limited sample are statistically indistinguishable. The coefficient for GNI in the IDA allocation model remains statistically insignificant at conventional levels, and the coefficient on CPIA rounds to .34 for both time periods. Similarly, the proxy for the MCC scorecard remains virtually unchanged. The coefficient on the WGI index rounds increases slightly from .377 to .388, which is within on standard error. The coefficient on corruption remains negative but statistically insignificant at conventional levels. The coefficient on democracy round to .10 for both samples and is statistically significant at the 1% level for both as well. This suggests that the findings in Table 5 are robust to the more limited sample of the aid effectiveness era from 2005 to 2020, and there wasn't a significant shift in the drivers of project outcomes in the more recent period.

Table A5. Fixed Effects Model of Project Outcome Ratings Regressed on Explanatory Variables in Different Period

	Full Sample (1995-2027)						
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
GNI per capita (ln)	0.1178***					0.0358	
	(0.0117)					(0.0239)	
CPIA overall score		0.3158***				0.3418***	
		(0.0308)				(0.0371)	
WGI index			0.2564***				0.3765***
			(0.0187)				(0.0353)
Corruption				0.2067***			-0.0036
				(0.0185)			(0.0327)
Democracy					0.0153		-0.0969***
					(0.0093)		(0.0115)
Observations	10,952	5,954	13,692	13,670	14,111	4,656	13,659

	Limited Sample (2006-2020)						
	(8)	**(9)**	**(10)**	**(11)**	**(12)**	**(13)**	**(14)**
GNI per capita (ln)	0.1025***					0.0365	
	(0.0141)					(0.0246)	
CPIA overall score		0.3113***				0.3370***	
		(0.0315)				(0.0380)	
WGI index			0.2256***				0.3880***
			(0.0227)				(0.0425)
Corruption				0.1616***			-0.0518
				(0.0230)			(0.0413)
Democracy					-0.0018		-0.1020***
					(0.0115)		(0.0141)
Observations	6,716	5,441	8,023	8,028	8,043	4,313	8,023

Note: OLS regression of project outcomes on various allocation criteria with fixed effects for different time period
* p < .1, ** p < .05, *** p < .01.

Source: PPD v2.0, WDI, WGI, and Freedom House.

Appendix 4. Testing robustness to different sets of fixed effects

In this appendix, I test whether including different sets of fixed effects or their timing changes my results. In my preferred specification, I include donor, sector, and year fixed effects. I include that specification here, but I also test whether only including donor and sector fixed effects or just donor fixed effects make an appreciable difference in the results. My results are presented in Tables A6 through A9.

Table A6 shows the specifications for GNI per capita, CPIA scores, and the WGI index. The estimates in column (1) are the same as in Table 5 above. The coefficients remain virtually unchanged when shifting to columns (2) and (3) that no longer includes year fixed effects and sector fixed effects, respectively. All of these coefficients remain statistically significant at the 1% level.

Table A6. Fixed Effects Model of Project Outcome Ratings Regressed on Average Income with Different Fixed Effects

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
		Need			CPIA			WGI	
GNI per capita (ln)	0.1178***	0.1155***	0.1115***						
	(0.0117)	(0.0116)	(0.0114)						
CPIA overall score				0.3158***	0.3037***	0.2686***			
				(0.0308)	(0.0307)	(0.0300)			
WGI index							0.2564***	0.2501***	0.2253***
							(0.0187)	(0.0186)	(0.0181)
Donor Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓	✓
Sector Fixed Effects	✓	✓		✓	✓		✓	✓	
Year Fixed Effects	✓			✓			✓		
Observations	10,952	10,952	11,171	5,954	5,954	6,137	13,692	13,692	13,933

Note: OLS regression of project outcomes on various allocation criteria with fixed effects for projects complete 1995 or later.
* p < .1, ** p < .05, *** p < .01.
Source: PPD v2.0, WDI, and WGI.

There are greater differences in the coefficients for the CPIA scores. When not using year fixed effects in column (5), the coefficient decreases from .316 to .304. This decreases further when sector fixed effects are not included in column (6) to .269. The coefficients all positive and statistically significant at the 1%

level, however. The coefficients for the WGI index also decrease but not as much. The coefficient decreases from .256 in column (7) to .250 in column (8) when year fixed effects are removed. Similarly, the coefficient decreases further in column (9) when sector fixed effects are removed. This is only about a 10% decrease from column (7) to column (9), however, and the coefficients all remain positive and statistically significant at the 1% level.

Table A7 shows the specifications for the control of corruption indicator from WGI and the democracy index from Freedom House. The estimates in column (1) and (4) are the same as in Table 5 above. The coefficients remain virtually unchanged when shifting from columns (1) to column (2) for corruption. The coefficient decreases slightly when only including donor fixed effects, however, though not by much. All of the coefficients remain positive and statistically significant at the 1% level.

Table A7. Fixed Effects Model of Project Outcome Ratings Regressed on MCC Hard Hurdles with Different Fixed Effects

	(1)	(2)	(3)	(4)	(5)	(6)
	Corruption			Democracy		
Corruption	0.2067***	0.2028***	0.1865***			
	(0.0185)	(0.0184)	(0.0181)			
Democracy				0.0153	0.0166*	0.0073
				(0.0093)	(0.0093)	(0.0092)
Donor Fixed Effects	✓	✓	✓	✓	✓	✓
Sector Fixed Effects	✓	✓		✓	✓	
Year Fixed Effects	✓			✓		
Observations	13,670	13,670	13,911	14,111	14,111	14,352

Note: OLS regression of project outcomes on various allocation criteria with fixed effects for projects complete 1995 or later.
* p < .1, ** p < .05, *** p < .01.
Source: PPD v2.0, WGI, and Freedom House.

The coefficient for democracy changes a bit more. The coefficient in column (4) from Table 5 above was statistically insignificant, but this becomes statistically significant at the minimal 10% level when year fixed effects are not included. The coefficient remains small at just .017, however, which means that a one standard deviation increase in democracy would lead to a .017 increase in a country's average project outcome rating. This significant relationship does not hold up when only including donor fixed effects, however. This suggests that the small coefficient on democracy is not particularly robust.

Table A8. Fixed Effects Model of Project Outcome Ratings Regressed on Donor Allocation Modelswith Different Fixed Effects

	(1)	(2)	(3)	(4)	(5)	(6)
		IDA			MCC	
GNI per capita (ln)	0.0358	0.0430*	0.0593**			
	(0.0239)	(0.0238)	(0.0232)			
CPIA overall score	0.3418***	0.3226***	0.2822***			
	(0.0371)	(0.0369)	(0.0357)			
WGI index				0.3765***	0.3636***	0.3378***
				(0.0353)	(0.0352)	(0.0346)
Corruption				-0.0036	-0.0032	0.0038
				(0.0327)	(0.0327)	(0.0324)
Democracy				-0.0969***	-0.0925***	-0.0977***
				(0.0115)	(0.0114)	(0.0113)
Donor Fixed Effects	✓	✓	✓	✓	✓	✓
Sector Fixed Effects	✓	✓		✓	✓	
Year Fixed Effects	✓			✓		
Observations	4,656	4,656	4,823	13,659	13,659	13,900

Note: OLS regression of project outcomes on various allocation criteria with fixed effects for projects complete 1995 or later.
* p < .1, ** p < .05, *** p < .01.
Source: PPD v2.0, WDI, WGI, and Freedom House.

Table A8 examines the robustness of the estimates for the donor allocation models. Columns (1) and (4) are the same specifications as in Table 5 in the main text that includes the full set of fixed effects, i.e., donor, sector and year. When different combinations of fixed effects are inluded, the results for the IDA allocation model change slightly, but the proxy for the MCC scorecard does not. While the coefficient for GNI per capita is not statistically significant in the preferred specification for the IDA allocation model, it is statistically significant at the 1% level in column (2) with donor and sector fixed effects and slightly larger. This trend continues in column (3) that includes only donor fixed effects. The coefficient on GNI per capita is slightly larger and statistically significant at the 5% level. The oppositive is shown for the CPIA score coefficient. The coefficient gets small as I move to column (2) and (3). The interpretation is not significantly different, however – CPIA is a significant driver of project outcomes in that better-governed countries perform better, whereas average income plays less of a role in that richer countries do slightly better in terms of project outcomes, on average.

For the MCC scorecard proxy in columns (4) through (6), ther are not large changes in the coefficients. The coefficient on the WGI index decreases slightly from .377 to .364 in column (5) and .338 in column (6). This is roughly within one standard error of the estimate with the full set of fixed effects, hover. The coefficient on corruption gets more positive when moving from column (4) to (6) and taking away the year and sector fixed effects, but the estimates remain statistically insignificant at conventional levels. Finally, the coefficient on democracy remains virtually unchanged from -.097 in column (4) to -.093 in column (5) to -.098 in column (6). All three estimates are statistically significant at the 1% level. These specifications confirm the importance of good governance in WGI, the insignificance of the corruption indicator, and that democracies tend to do worse in terms of project outcomes.

Finally, I examine whether the timing of the time fixed effects matter. This addresses the inconsistency of the timing of the explanatory variables and the year fixed effects. That is, the explanatory variables are averaged over the life of the project, whereas the year fixed effects in my preferred specification are from the completion year of the project. Table A9 shows the estimates for the IDA and MCC allocation models when I use a different year for the time fixed effects. In columns (2) and (4) I use the mid-point of the project, which is just the halfway point of start year and completion year (rounded up). In columns (3) and (6), I use the start year of the project. The estimates are nearly identical to the preferred specification using the completion year for the time fixed effects, so I am not concerned about using the completion year instead of one of these other years.

Table A9. Fixed Effects Model of Project Outcome Ratings Regressed on Donor Allocation Modelswith Different Fixed Effects

	(1)	(2)	(3)	(4)	(5)	(6)
		IDA			MCC	
GNI per capita (ln)	0.0358	0.0305	0.0334			
	(0.0239)	(0.0239)	(0.0238)			
CPIA overall score	0.3418***	0.3344***	0.3277***			
	(0.0371)	(0.0372)	(0.0372)			
WGI index				0.3765***	0.3851***	0.3841***
				(0.0353)	(0.0354)	(0.0354)
Corruption				-0.0036	-0.0046	-0.0036
				(0.0327)	(0.0327)	(0.0326)
Democracy				-0.0969***	-0.1014***	-0.1012***
				(0.0115)	(0.0115)	(0.0115)
Donor Fixed Effects	✓	✓	✓	✓	✓	✓
Sector Fixed Effects	✓	✓	✓	✓	✓	✓
Year Fixed Effects (end year)	✓			✓		
Year Fixed Effects (mid year)		✓			✓	
Year Fixed Effects (start year)			✓			✓
Observations	4,656	4,656	4,656	13,659	13,659	13,659

Note: OLS regression of project outcomes on various allocation criteria with fixed effects for projects complete 1995-2027.
* p < .1, ** p < .05, *** p < .01.
Source: PPD v2.0, WDI, WGI, and Freedom House.

Appendix 5. Robustness check of whether donors grade to a curve

One potential concern is related to the comparability of the ratings for individual donors, though the previous literature has found more commonalities than differences in trends across institutions. Bulman *et al* (2016) compare correlates of project outcomes in World Bank versus Asian Development Bank projects – two of the largest multilateral donors. They find similar results across the two institutions and cannot reject the null that the relationship is the same across the two institutions. Similarly, Briggs (2020) tests the importance of economic growth and democracy in achieving project outcomes using an earlier version of the dataset utilized here. That research has more of a focus on comparability of results across donors, and he finds that his results are reasonably similar across donors and are thus generalizable to other donors.

Nonetheless, there could be important differences in donor rating systems. While many donors in the sample claim that they objectively assess their projects or even have an independent evaluation unit, it is possible that some donors "grade to a curve" while other do not. That is, certain countries are inherently more difficult environments in which to work, and therefore, a project outcome in a more difficult country might be given a higher rating than a comparable project outcome in a more accommodating operating environment. To be clear, I am not claiming that donors are giving themselves better assessments to look good, rather, it is possible that this happens without the donor even realizing it. If a donor is used to repeated failure in a country or has low ambitions because a country is very poorly governed or conflict-affected, then they may unknowingly provide a higher rating than than they might have otherwise.

To test whether this is the case, I construct a test across donors that examines their average ratings in the most difficult and most conducive countries and compare them to their overall average rating. I do this by defining two new groups – one with the highest income and best governance and a second with the lowest incomes and the worst governance. This is a reasonable proxy for the most difficult and easiest operating environments that donors might face.

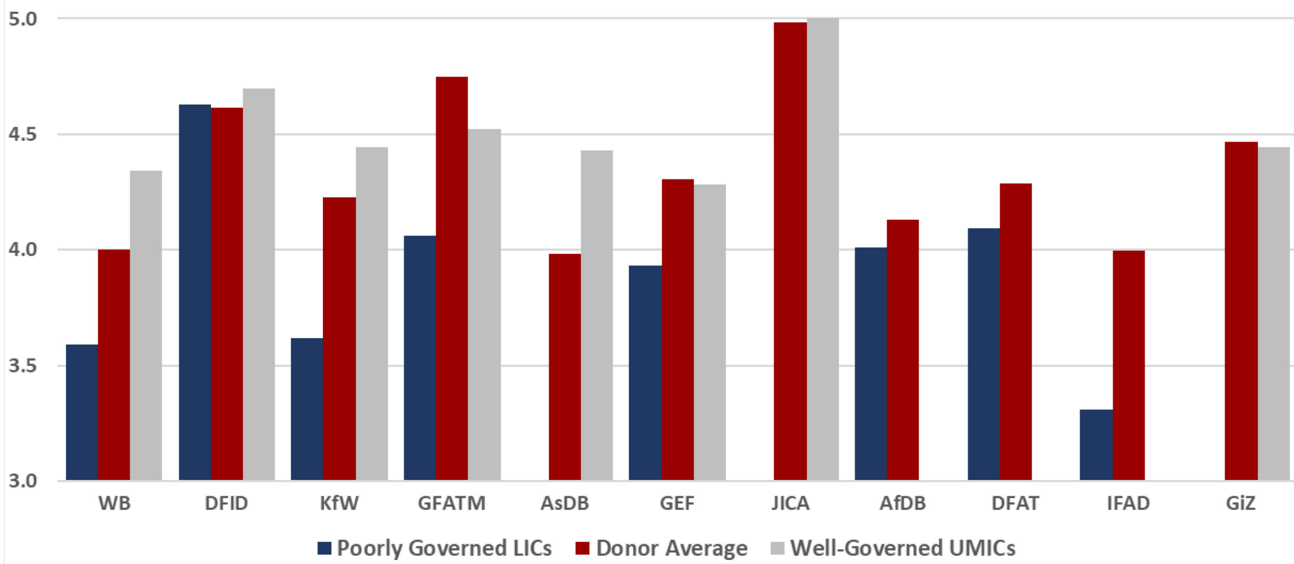
More specificially, my criteria are (1) average income; (2) the WGI indicators on the MCC scorecard; and (3) the number of projects a donor has completed in each group. First, I create an average of each country's GNI per capita over the entire sample, and I only include countries that are above the LMIC/UMIC threshold of $4,255 in FY 2023. This implies that the included countries were probably a UMIC for most of the sample and/or graduated out of the sample at some point by becoming a HIC. Second, I only included the best-governed countries by only including countries that were more than one standard deviation above the mean for the index of MCC-relevant WGI indicators (i.e., regulatory quality, rule of law, and government effectiveness) for their average across the entire sample. This implies that they were very well-governed for most, if not all, of the entire sample period. Finally, I only included countries that 30 or more projects for all donors. This narrowed the group to just 12 countries.[53]

For the other end of the spectrum, I did conduct a similar process to arrive at a small group of the most difficult operating environments. First, I only included countries that were below the LIC/LMIC threshold for FY 2023 of $1,085 or less in terms of average GNI per capita over the entire period. This implies that the country was a LIC for most of the period, if not all of it. Second, I only included countries whose average governance score for the full time period was more than one standard deviation below the mean of the WGI index for the full sample. This means that only the consistently worst-governed

[53] These 12 countries are: Bulgaria, Chile, Costa Rica, Croatia, Malaysia, Mauritius, Panama, Poland, South Africa, Thailand, Turkey, and Uruguay.

coutnries were included. Finally, I again excluded any countreis with fewer than 30 projects across all donors. This produced a group of just 10 countries.[54]

Figure A2. Average Ratings by Donor for Various Country Groupings



Source: Project Performance Database 2.0.
Note: Average rating by donor with more than 10 projects for (1) countries that are LICs that are at least one standard deviation below the WGI mean; (2) all projects for that donor; and (3) UMICs at least one standard deviation above the WGI mean.

I then compared each donor's average rating for each group to its full sample average, while excluding any donor with fewer than ten projects for a given group. The results are of this analysis is presented in Figure A2. There is a pretty clear trend overall – the poorly governed LICs score well below the mean, while the well-governed UMICs score better than the average for all projects. The average for all projects is 4.24, while the average for poorly governed LICs is 4.03 and 4.47 for well-governed UMICs. This pattern holds up for almost every single donor, which are displayed in order of their total number of projects (left to right) with the World Bank having the most. The World Bank shows a particularly sharp difference between these groups with about a 3.6 for the poorly governed LICs and a 4.34 average for the well-governed UMICs relative to their 4.0 overall average. One noticeable discrepancy from the

[54] These ten countries are: Afghanistan, Burundi, Central African Republic, Chad, Democratic Republic of the Congo, Guinea, Guinea-Bissau, Liberia, Sierra Leone, and Somalia.

expected pattern is DFID whose poorly governed LICs score just above the full sample average, though their well-governed UMICs are above their average rating. Also, JICA and GiZ did not have significantly better scores for their well-governed UMICs as they are both very close to their overall average. These are exceptions to the rule, however, as almost every other donor has lower scores for the poorly governed LICs and higher scores for the well-governed UMICs. This suggests that most donors are not "grading to a curve."

Nonethless, it is worth test whether these donors make a difference in affecting my results, and I exclude those three donors (DFID, JICA, and GiZ) in my preferred sprecifications for the IDA and MCC allocation model with the full set of donor, sector, and year fixed effects.

The results are show in Table A10. The main specification shown in previous regressions is shown in columns (1) and (3). The donors that may be grading to a curve are show in columns (2) and (4) and all other donors are show in columns (3) and (6). If these donors were indeed grading to a curve, we would see insignificant relationships to the various of interest. This is not the case, however. For the IDA allocation model, DFID, JICA, and GiZ have a very similar set of results to the baseline, though the coefficient is smaller for CPIA. It is still statistically significant, however, which means that governance as measured by CPIA scores does have an effect on those three donors project outcomes.

A similar pattern emerges for the MCC allocation model specifications. The coefficients for the WGI index are not statistically different from each other, the control of corruption indicators is statistically insignificant at conventional levels, and the democracy index has a negative coefficient that is statistically significant for all three groups at the one percent level. The demoracy coefficient for DFIC, JICA, and GiZ is actually larger in absolute terms, which suggests that democracy has a greater negative effect on project outcomes.

While the results are slightly different for DFID, JICA, and GiZ, these estimates do not suggest that DFID, JICA, and GiZ are grading to a curve. Rather, their projects have the same general relationships with the allocation criteria, and in one case (democracy), an even stronger relationship than the other donors. Importantly, the main results also hold up when not including these three donors.

Table A10. Fixed Effects Model of Project Outcome Ratings Regressed on Donor Allocation Modelswith Different Fixed Effects

	(1)	(2)	(3)	(4)	(5)	(6)
	IDA			MCC		
	All Donors	**DFID, JICA, GiZ**	**All Other Donors**	**All Donors**	**DFID, JICA, GiZ**	**All Other Donors**
GNI per capita (ln)	0.0358	-0.0703	0.0515*			
	(0.0239)	(0.0511)	(0.0274)			
CPIA overall score	0.3418***	0.2240***	0.3565***			
	(0.0371)	(0.0788)	(0.0429)			
WGI index				0.3765***	0.3683***	0.3997***
				(0.0353)	(0.0863)	(0.0391)
Corruption				-0.0036	-0.0803	0.0076
				(0.0327)	(0.0841)	(0.0356)
Democracy				-0.0969***	-0.1233***	-0.0944***
				(0.0115)	(0.0270)	(0.0127)
Donor Fixed Effects	✓	✓	✓	✓	✓	✓
Sector Fixed Effects	✓	✓	✓	✓	✓	✓
Year Fixed Effects (end year)	✓	✓	✓	✓	✓	✓
Observations	4,656	905	3,751	13,659	2,613	11,046

Note: OLS regression of project outcomes on various criteria with fixed effects for different donors for projects completed 1995 or later.
* p < .1, ** p < .05, *** p < .01.
Source: PPD v2.0, WDI, WGI, and Freedom House.

Chapter 5. Optimal Allocation Models

This concluding chapter explores the current allocation of assistance and compares it to an optimal allocation of assistance. This chapter contributes to the normative strand of the related literature that provides prescriptive models and recommendations for improving the allocation of assistance in pursuit of large potential effectiveness gains. For instance, Collier and Dollar (2002) proposed a "poverty efficient" allocation that would have the greatest impact on poverty and found that a better allocation could roughly double aid's poverty impact. Similarly, Bigsten *et al* (2011) found that approximately €19 billion of the EU's €27 billion in aid should be re-allocated with a potential net gain of about €7.8 billion when more explicitly accounting for the quality of governance.[55] These large shifts are because donors either do not have allocation criteria in place (Ottersen 2017) or the actual distribution of aid does not resemble the criteria in practice (Ceriani and Verme 2014). This chapter synthesizes and employs my previous findings by applying the evidence regarding the various allocation criteria produced in the previous chapters to recommend a way forward to a more efficient allocation of assistance.

My analytical approach is to estimate multiple "optimal" allocations of assistance and then compare them to the current allocation of assistance.[56] I start by describing the current allocation of assistance. Easterly and Pfutze (2008) described an "ideal" aid agency and developed a measure of selectivity to summarize how much aid goes to poor countries, i.e., the proportion of aid going to LICs. While it possible to devise a more sophisticated measure, such as the weighted average of recipient income used

[55] Using similar assumptions as Collier and Dollar (2002), the authors found that the poverty reducing effect of aid to over-aided countries only had about 15 percent of the effect that the same resources would have in under-aided countries, i.e., the gain would be 85 percent of the mis-allocated 19 billion. However, they also assume that aid must be adjusted for the quality of governance in achieving aid effectiveness because governance is not as good in the under-aided countries, and this reduces the net gain by about half to 7.8 billion. In essence, the authors were claiming that the same poverty impact could be achieved with far fewer resources.

[56] I deliberately use the term "optimal" incorrectly (as there can only be one optimum), as this is the term used in the literature. As I will show, this is particularly ironic because there is no "optimal" allocation of criteria, so these would be more accurately described as "prescriptive models".

to assess UK aid by Hughes and Mitchell (2023),[57] I start by following Dissanayake *et al* (2020) in describing the distribution of aid across income groups. This gives a sense of how much a donor is favoring the poorest countries versus providing assistance to UMICs, and it is particularly useful in comparing across donors. Dissanayake *et al* (2020) found that a significant proportion of assistance was flowing to UMICs, and I confirm this observation.

I then proceed to comparing the actual allocation of aid resources to "optimal" allocation models that incorporate findings from my previous chapters. Beyond the Collier and Dollar (2000) poverty-efficient approach, McGillivray (2003a) surveyed the various prescriptions for how aid *ought* to be allocated and Ottersen *et al* (2017) simulated 11 different allocation criteria (but confine their analysis to global health assistance). I take a similar approach and set up the analysis by producing five different "optimal" allocations of assistance by specifying three key parameters: need, effectiveness, and population. I start with two performance-based approaches: (1) the IDA allocation formula (of average income, policy performance, and population) extended up to the IBRD line; and (2) a variant on the IDA allocation formula proposed by DFID (Dercon and Lea 2016) that elevates the relative importance of average income. Mitchell and Hughes (2020) only include LICs and LMICs in their IDA model, but I incorporate the findings from my country need chapter to extend the models to relatively poor UMICs (while not allocating assistance to relatively richer UMICs). I apply this threshold to all five optimal allocation models. Recall my earlier findings that the effectiveness of good governance was split along the micro-macro divide – better governance does not improve the aid-growth relationship at the macro level, but it is critically important for project success at the micro level. This suggests that a step back from the

[57] Hughes and Mitchell (2023) analyzed the shift in UK aid away from a needs-based approach, so they took a more granular approach that also captured shifts within income groups towards less-poor countries. While this is arguably a more appropriate measure for analyzing changes over time, I am more concerned with a snapshot of donor allocations, and the distribution across income groups provides a broad sense of how needs-based a donor's allocation may be.

heavy emphasis of policy performance in the IDA model (over need) may be warranted. The DFID model does this by emphasizing need over policy performance by changing the relative weights of the parameters. However, it is arguable whether a performance-based approach should be employed at all.

To test a need-based approach, I also produce three models that set aside policy performance: (3) the same need and population parameters as DFID (that excludes effectiveness); (4) a country's share of the global number of people living in extreme poverty; and (5) a forward-looking measure of the projected number of person-poverty-years expected between 2019 and 2030 as a proportion of the global total. Given the strong correlation between average income and poverty, allocation models (3) and (4) are not that different. Barder (2009) observed that current misallocations may be partially explained by donor expectations of poverty rates at some point in the future and proposed that allocation models account for expected future poverty to be more in line with these preferences. The forward-looking allocation model (5) exploits poverty projections produced by Cuaresma *et al* (2018) to anticipate where extreme poverty is likely to persist. This a useful proxy for donor expectations of where the development challenge is likely to persist, which also partly reflects understandings of which countries are self-sufficient and/or already on the path to ending extreme poverty. In building out these various allocation models, my approach builds on Mitchell and Hughes (2020) that explore four similar allocation models but confine their analysis to global assistance – I extend their analysis by dis-aggregating aid among various donor groups with a focus on the United States as the largest bilateral donor. Once I produce estimates using the various allocation models, I then compare current allocations to optimal allocations on a country-by-country basis to see where the largest discrepancies lie.

In terms of current allocations, I find that about half of total official assistance goes to LMICs, and LICs and UMICs receive roughly one quarter each. I find that both multilateral and bilateral donors allocate about half of their assistance to LMICs, but multilateral donors allocate more to LICs and less to UMICs and bilateral donors do the opposite – they allocate more to UMICs and less to LICs. Turning to the optimal allocation models, about 90 percent of total assistance is evenly split across LICs and LMICs, except in the IDA allocation model. This is because the IDA optimal allocation model does not reduce allocations significantly for average income and instead emphasizes policy performance – as a result, allocations to LICs are substantially lower (and allocations to UMICs are substantially higher) than the other four models. The DFID allocation model emphasizes need over effectiveness and is more in line with the needs-based approaches, which all approximately evenly distribute aid between LICs and LMICs. The allocation models based on a country's current and future share of the global poor are very similar to the GNI model. This is not surprising given the strong correlation between average income and poverty measures that I show in the second chapter. As might be expected, global poverty is projected to persist primarily in the poorest countries when we look ahead to 2030.

At a country level, the largest recipient of official assistance is Bangladesh followed closely by India. Pakistan and Ethiopia are the second and third largest multilateral recipients. While there are no UMICs on the multilateral list, Indonesia and Jordan are the third and fifth largest DAC aid recipients, respectively, and Jordan is the top recipient of US assistance. The optimal allocation models produce very different top recipients, however. For instance, the top country from the IDA allocation model is Indonesia, a well-governed UMIC, that is not among the top ten multilateral or US recipients of assistance. This points to the weakness of the IDA allocation model – it strongly favors better-governed countries with only a slight penalty for higher average incomes. The DFID allocation model more closely resembles the actual multilateral donor allocation. It tends to favor large-population poor countries, such as Ethiopia, DRC, and India. The GNI model (without the effectiveness parameter) produces a similar list, though some of the poorly governed countries move up (DRC and Nigeria), while some better-governed countries move down (Indonesia and Uganda). Naturally, the extreme poverty model reflects where the greatest number of poor currently live and is largely populated by African countries. The future poverty model has even more of an African bias with the only non-African country among the top ten recipients being India (sixth). This is in line with projections that the future of global poverty will overwhelmingly be in Africa (World Bank 2022).

Finally, I compare current allocations to optimal allocations on a country-by-country basis to identify the most over- and under-aided countries. For total official aid, the IDA model shows that Afghanistan and Jordan are the most over-aided countries followed by several other conflict-affected states (e.g., Somalia and Yemen). There is significant overlap between the IDA formula and the other formulas, which are mainly comprised of strategic partners in the Middle East, such as Egypt, Jordan, and Iraq. The under-aided countries from the IDA model are disproportionately large-population UMICs, such as Indonesia. The DFID model puts Ethiopia and Pakistan at the top with eight of the ten under-aided countries being African states. DRC and Nigeria are at the top of the needs-based allocation models, and most of the other under-aided countries are also in sub-Saharan Africa. Indeed, all ten under-aided countries from the future poverty allocation model are in Africa.

For US assistance, there is more unanimity across the allocation models in terms of over-aided countries. Jordan is the most over-aided country for every single allocation model by $830 to $877 million. Afghanistan is in the top three on all lists except current extreme poverty, and South Africa is also among the top six countries for all five models. Colombia and Iraq are also considered highly over-aided for all models except IDA. The most under-aided countries differ much more by the allocation model than the over-aided countries for the US, and the under-aided countries more closely resemble that of official aid. Indonesia, Pakistan, and India are the most under-aided countries for US assistance according to the IDA model. Ethiopia and Pakistan top the DFID model's list of under-aided countries. The top three countries for the GNI, extreme poverty, and projected poverty models are also almost exactly the same – DRC and Pakistan are the most under-aided in terms of GNI, and DRC and Nigeria are the most under-aided when it comes to both current and future poverty.

My contributions in this chapter are two-fold. First, I extend the analysis of Mitchell and Hughes (2020) beyond total official aid to analyze (and criticize) the bilateral allocations of the United States in particular. By disaggregating total aid, I am able to separately examine the allocations of multilateral donors, the traditional bilateral donors in the DAC, and the largest bilateral donor in the US. Similarly, McGillivray (2003a) surveyed the various prescriptions for how aid should be allocated and compared the models to the DAC and Sweden.[58] In particularly, I find that the US is heavily over-invested in UMICs and strategic partners in the Middle East. This dis-aggregated analysis provides additional context and more actionable findings for donors beyond characterizing the highly fragmented aid system as a whole. Second, throughout my analysis I explore the implications of the parameters of each of the optimal allocation models. In general, the IDA model favors the best-governed countries regardless of average income, the DFID model favors the best-governed poor countries, the GNI model favors the world's poorest countries, the extreme poverty model favors countries with the highest poverty headcount ratios, and the projected poverty model favors countries with the highest projected poverty rates in the future. While Mitchell and Hughes (2020) produced a list of over-aided countries in addition to the under-aided countries previously identified by the OECD (2013), they do not dig into the reasons why their models produced different allocations and why these differ from current allocations. In that way, my analysis informs policymakers on which allocation model might make the most sense for their

[58] Twenty years ago, McGillivray (2003a) cited Sweden being "of special interest, as it usually rates highly in aid allocation evaluations," and this remains the case today – Sweden remains the top country with a 100 score in CGD's Commitment to Development Index: https://www.cgdev.org/cdi. For perspective, France is second at 78.

circumstances. That is, the evidence produced in this chapter can help to inform policy choices regarding the most appropriate allocation model and once decided, also provides a list of the highest priority countries for which to increase or decrease assistance. My analysis and recommendations focus mainly on the US aid system, but these insights could easily be extended to other donors.

I conclude with policy recommendations. First, most donors would benefit from more explicitly stating their allocation criteria. While there is no obvious optimal allocation model, stating a set of allocation principles, if not explicit criteria, is helpful for both transparency and discipline. The question then becomes which model is best, and this chapter provides insights needed to make a more informed decision. Once a policymaker understands the trade-offs and implications of each potential allocation model, they can make an evidence-based choice of the model that best suits their preferences and constraints. Second, donors should establish a more explicit threshold for graduation from assistance. More formal rules around graduation would be useful, at least partially because it forces a donor to identify its allocation criteria and make a determination of when they believe aid is no longer needed.

Data

I use a combination of official aid data and country-level factors. I examine the current allocation of assistance using data from the OECD's database, *OECD.stat*. I use the "country programmable aid" (CPA) data, which isolates a subset of ODA flows that are multi-year programs at the country level, averaged over the period of 2019 through 2021 to smooth out any volatility in year-to-year aid flows. CPA data exclude humanitarian assistance and debt relief that would likely not fall under the goal of poverty reduction or economic growth. Conceptually, CPA flows are assistance that leaves the donor country and are negotiated with the partner country government, i.e., it is not spent on administrative costs or within-donor refugee hosting costs. Furthermore, these data do not include food aid, NGO funding, equity, or flows that cannot be attributed to a particular country. Finally, CPA data are a gross measure that does not account for loan repayments, as these are not considered in allocation decisions. All of these characteristics make CPA particularly useful for examining donor allocation policy.

I focus mainly on total official and US aid flows. For comparison purposes, I also reference total multilateral assistance as well as DAC assistance as a benchmark for bilateral donors. These benchmarks include IDA (multilateral) and the US (DAC), and the official total includes all of these dis-aggregates. Between the multilateral and DAC figures, I cover almost all of official assistance, though this does not include non-DAC donors, like the Gulf States and China.[59] For the allocation criteria, I obtain and merge data from several different sources. I utilize GNI per capita (Atlas) as my measure of average income and a population estimate from the World Bank's World Development Indicators (WDI, which is averaged over the period of 2017 through 2019. The Country Performance Rating (CPR) is taken directly from the World Bank's IDA allocation documentation for 2019.[60] I impute this rating using World Governance Indicators (WGI) scores for countries not listed in the official CPR scores or that are no longer IDA-eligible. For all of these indicators except the CPR and country income groups from 2019, I use an average over the period 2017 through 2019 to increase the sample size for countries missing 2019 data.

For the number of people living in extreme poverty, I use 2019 estimates from the World Bank's Poverty and Inequality (PIP) for the $2.15 per day poverty line. For poverty projections, I use the Cuaresma (2018) replication file, which is the study that created the original data for the World Poverty Clock,

---

[59] Over the period 2019-2021, DAC donors account for $53,582 billion and multilateral bodies account for $50,456 of the $109,620 of official country programmable aid. Combined, this is over 95 percent of the total. However, non-DAC donors are almost certainly under-counted in official OECD statistics as many of them are not transparent in sharing the volume of their aid flows. For instance, Grover (2023) used AidData's Tracking Underreported Financial Flows database to show that China committed $6.5 to $7.4 billion in "ODA-like" flows each year between 2014 and 2017. That amount alone is greater than total non-DAC donors reported by the OECD's data.

[60] The historical data for CPR scores are available here: https://ida.worldbank.org/en/financing/resource-management/ida-country-performance-ratings.

though those data have been updated to account for more recent development.[61] I use their Shared Socioeconomic Pathways scenario 3 (SSP3) that seems the most realistic and is most closely aligned with the current 2030 projection of the World Poverty Clock. Following Mitchell and Hughes (2020), I use the projected number of poor over the period 2019 through 2030 to calculate a share of the extreme poverty "person years" for each country. That is, each person living in extreme poverty in a country is counted for each year they are expected to remain in poverty. This figure is summed across the entire period by country and compared to the global projection to produce that country's share of expected poverty over the entire period.

Table 1. Summary statistics for Aid and Allocation Criteria

	Unit	Obs.	Mean	Std Dev	Min	Median	Max	Sum	Period	Source
GNI per capita (Atlas)	Current Dollars	129	3,902.4	3,013.4	240.0	3,250.0	12,263.3	N/A	2017-19	WDI
Population	Millions	128	37.1	126.0	0.0	9.9	1,350.0	6,147.4	2017-19	WDI
CPR Score	1-6 Scale	127	3.2	0.4	1.9	3.2	4.2	N/A	2019	WB
Extreme Poor	People (Millions)	120	5.4	16.7	0.0	0.4	160.0	657.0	2019	PIP
Future Poverty Years	Person Years (Millions)	123	47.4	136.0	0.0	5.1	1,210.0	5,858.7	2019-30	WPC
Official CPA	2020 Dollars (Millions)	128	856.4	989.7	25.1	497.1	5,189.0	110,377.3	2019-21	OECD
Multilateral CPA	2020 Dollars (Millions)	128	394.2	497.2	8.9	213.3	2,529.0	50,538.2	2019-21	OECD
DAC CPA	2020 Dollars (Millions)	128	418.6	566.1	2.6	202.4	3,976.7	54,208.3	2019-21	OECD
US CPA	2020 Dollars (Millions)	128	96.4	153.3	0.0	37.8	876.9	12,364.2	2019-21	OECD

Source: World Bank (WB); WDI; PIP; Cuaresma 2018; OECD.

Table 1 shows summary statistics for my data. I have between 120 and 129 observations for all of these indicators, which covers almost all developing countries (LICs, LMICs, and UMICs but not HICs). The mean of average income is that of a relatively rich LMIC at around $3,900 with a low of $240 (Burundi) and a high of $12,263 (Argentina). The average CPR score is 3.2 with a low of 1.88 (Somalia) and a high of 4.15 (Samoa). The average country has about 5.4 million people living in extreme poverty with Nigeria (61 million) and India (160 million) having the largest numbers. The poverty picture changes when

[61] The World Poverty Clock provides a real-time estimate of the number of people living in extreme poverty and tracker of progress towards SDG1: https://worldpoverty.io/headline.

shifting to future poverty years, however, where DRC and Nigeria have the largest projected number of person years over the period 2019 through 2030. The CPA data show that the largest recipient of total official and multilateral CPA is Bangladesh ($5.19 billion and $2.5 billion, respectively), India is the largest recipient of DAC CPA ($4.0 billion), and Jordan and Afghanistan are the largest recipients of US CPA ($876 million and $847 million, respectively).

Regarding timing, I use data for the allocation criteria from the period of 2017 through 2019 for most variables so that it is lagged from the actual allocation of aid commitments that are taken from the period of 2019 through 2021. This follows the aid allocation literature that observes both a data lag in official statistics as well as additional time that it takes for allocation policy decisions to translate into budget allocations and commitments (Grover 2009). The timing is not as critical of a concern as for studies determining aid allocation criteria – the allocation criteria used here are relatively sticky in that average income and governance scores move slowly as opposed to other potential determinants such as strategic interests that might shift quickly – and I am looking at the current allocation of criteria, not an extended time period. Nonetheless, I use lagged data for the allocation criteria (except the forward-looking projected poverty years) and the most current data available for actual aid commitments.

Analytical approach

The OECD (2013) frames the allocation criteria debate around three key parameters. First, how to conceptualize, measure, and weight development need. The OECD (2013) mentions the relevant debates regarding multi-dimensional poverty, but I showed in the second chapter that average income is the best concept and measure of development need – the only questions that remain are how much to weight that parameter and whether to impose a threshold at which aid is cut off. Nonetheless, while three of my five optimal allocation models include average income, I also include (current and future) poverty measures in two of the models. While Mitchell and Hughes (2020) choose to limit their IDA model to LICs and LMICs, I instead use the IBRD threshold given my finding in the country need chapter that $7,000 GNI per capita is an evidence-based threshold for cutting off assistance. This effectively allocates about one third of UMICs (19 of 58) zero assistance in the optimal models.[62]

Second, whether to include policy performance and how to weight it. The OECD (2013) references the extensive debates surrounding Burnside and Dollar (2000) and also mentions the views that prioritizing policy performance penalizes those most in need and that fragile states should be prioritized over strong policy performers. The question becomes how much to weight governance factors (if at all) relative to need. As a result, I employ a range of options to draw out the trade-offs associated with an effectiveness parameter. I utilize a high and a low relative weight for the effectiveness parameter in my performance-based models and exclude the effectiveness term entirely in the three needs-based allocation models.

Finally, how to account for differences in population size. The OECD (2013) poses this as a question of thinking in terms of aid per capita versus aggregate country envelopes, and they also observe that a small-country bias disadvantages countries with an overwhelming proportion of global poverty. In line with their “egalitarian” model, for three of my allocation models, I utilize a population parameter that would equalize aid per capita across otherwise-identical countries with different population size, all else equal. This implies no population bias in either direction. However, I also include two models based on the number of poor, which would be in line with the view of allocating assistance to poor people instead of poor countries, though I will show that there is minimal difference with a needs-based approach based solely on average income.

[62] This mostly binds for the IDA model that only slightly discriminates on the basis of average income, but it does force adjustments to all five optimal allocation models, though some are very minor.

With the framing around these parameters in mind, I now explain the details of the five different allocation models that will guide the rest of my analysis. My approach is to estimate these optimal allocations of assistance and then compare them to the current allocation of assistance, both for aid as a whole (all official donors) and the US specifically. The formula for each allocation model is below where *c* is the country observation, *y* is the year, and *d* is the donor:

(1) $\textit{IDA Allocation Score}_c = CPR_c^3 * Population_c * GNI/capita_c^{-0.125}$

(2) $\textit{DFID Allocation Score}_c = CPR_c^3 * Population_c / \textit{GNI per capita}_c$

(3) $\textit{GNI Allocation Score}_c = Population_c / \textit{GNI per capita}_c$

(4) $\textit{Extreme Poverty Share}_c = \frac{\textit{Extreme Poor}_c}{\textit{Extreme Poor (global)}}$

(5) $\textit{Future Poverty Years}_c = \frac{\sum_{t=1}^{T} \textit{Extreme Poor}_{ct}}{\sum_{t=1}^{T} \textit{Extreme Poor (global)}_t}$

(6) $\textit{Country Allocation Share}_c = \frac{\textit{Allocation Score}_c}{\sum_{c=1}^{C} \textit{Allocation Score}_c}$

(7) $\textit{Country Allocation}_{cd} = \textit{Allocation Share}_c * \textit{Resource Envelope}_d$

Equation 1 shows the IDA allocation model that uses GNI per capita to proxy need, CPR to proxy effectiveness, and population as its three key parameters. The CPR indicator varies across countries and has an exponent of three. The population indicator is a scaling variable with an exponent of one that varies across countries. This implies that there is no bias on per capita terms for small countries even though there is a small-country bias in most actual allocations.[63] GNI per capita varies by country and has an exponent of negative one eighth, which implies that allocations are gradually decreased as average income increases. The small negative exponent means that the need parameter only slightly discriminates on the basis of average income. For example, the term is the equivalent of .422 at $1,000 GNI per capita (LIC/LMIC line), .354 at $4,000 (LMIC/UMIC line), and .331 at $7,000 (IBRD line) – just more than a 20 percent decrease from the LIC/LMIC threshold to the IBRD line. At the extremes, the term gives .504 for the poorest country (Burundi) at $240 GNI per capita and .308 for the richest country (Argentina) at over $12,000 per capita. In contrast, the CPR term produces more than a ten-fold difference between the best-governed country (Samoa) at 71.5 (4.15^3) and the worst-governed country (Somalia) at 6.64 (1.88^3).

Equation 2 is a variant on the IDA allocation model proposed by DFID (Dercon and Lea 2016) that elevates the relative importance of need in the IDA allocation formula by changing the exponent on GNI per capita from -.125 to -1. In contrast to the small differences provided by the IDA formula, the DFID formula quickly decreases the allocation as a country grows in average income – decreasing by half the term each time a country doubles its average income. That is, the GNI/capita term for a country at $1,000 GNI per capita is .001, .0005 at $2,000, and .00025 at $4,000. At the extremes, the term is .0042 for Burundi and .00015 for the last UMIC before the IBRD threshold (Gabon) – more than a 25-fold difference. The other terms stay the same as the IDA allocation formula. Equation 3 is the same GNI per

[63] Anderson (2008) observes that population has a large, negative effect on aid per capita, and this is due to economies of scale; because smaller population countries are more vulnerable to shocks; and potentially due to international equity concerns related to allocating outsized assistance to a small number of large states. Anderson (2008) thus recommends reducing slightly the negative relationship with population, and I incorporate this into these models by eliminating the small-country bias in the optimal allocation models.

capita term as the DFID formula in addition to the population scaling term – it does not include the effectiveness term. This is a purer needs-based allocation approach that ramps down assistance for countries as they get richer and does not discriminate on the basis of policy performance. This is the first of three needs-based approaches.

Equation 4 is a country's share of the global number of people living in extreme poverty. This is a pure needs-based approach, as it does not include an effectiveness parameter. This does not additionally include a scaling variable for population, as this is already incorporated by using the number of people living in poverty, which is essentially the poverty rated scaled by the total population size. This approach suggests that a country's proportion of global aid should be dictated by its proportion of the global extreme poor – in other words, it very narrowly casts the development challenge as that of ending extreme poverty. Equation 5 is similar except that it is forward-looking in that it utilizes poverty projections to estimate the number of person poverty years expected between 2019 and 2030 as a proportion of the global total. The calculation takes the projected number of people living in extreme poverty in each country and sums that over the period of 2019 through 2030. This is compared the global total that is also summed across that period. In that way, it is similar to the current snapshot of extreme poverty in Equation 4 except that it builds in forward-looking expectations by donors of where extreme poverty is likely to persist.

Equations 1 through 3 require an extra step, as they do not produce an allocation share directly like Equations 4 and 5. Equation 6 shows that I must first sum the allocation scores across all countries for the denominator and then I can divide each country's allocation score by the sum to produce each country's allocation share of the relevant resource envelope. For all of these formulas, I then multiply the allocation share by the donor's total resource envelop to produce the country allocation for that pot of resources, as shown by Equation 7. This is the "optimal" allocation according to these various allocation models that can then be compared to the actual allocations of assistance. Alternatively, I could skip Equation 7 and compare each country's share of the total resource envelope to their optimal allocation share, but this only shows proportions of the total envelope instead of total resource amounts, which helps provide a sense of scale in terms of the misallocations. I do use allocation shares when comparing different donor groups, however, as it is necessary to keep these different pots of resources in the same scale so that they are comparable (see Table 2 below).

As a final step, I must make minor adjustments to the allocation shares produced (beyond zeroing out allocations for countries above the IBRD line as previously noted). Collier and Dollar (2002) note that aid allocation must be politically constrained by *ad hoc* caps on allocations to large-population countries, as their "poverty efficient" normative allocation model implies an overwhelming allocation of funding to India given its large number of poor and relatively strong policy environment. Following Mitchell and Hughes (2020), I cap India's allocation at 5 percent for all the models except for future poverty (where it does not exceed that). This is a reasonable decision, as India is currently at slightly less than five percent of global aid (4.7 percent) and has made it clear that they do not want more (Mitchell and Hughes 2020). I also choose to allocate China zero assistance in all models for related reasons – China has become a major donor in its own right and currently receives less than one percent of global assistance. Finally, while there is a significant literature on absorptive capacity,[64] I do not find this to be a major concern in implementing these models. When imposing a cap of 30 percent of GNI in PPP terms, I find that only

[64] For instance, Carter (2014) questions whether a dogmatic pursuit of greater donor concentration of resources in the poorest countries translates into greater effectiveness. The author argues that when countries run up against the limits of their aid absorption capacity, further allocations that favor those countries may not be optimal. This implies that the weight placed on the income criteria in aid allocation formulas may be less than existing models, and this would avoid too much aid being directed to poor countries that are unable to absorb it effectively.

one country was slightly above this absorptive capacity limit, Burundi.[65] Given that Burundi has the smallest GNI per capita in the sample and the aid to GNI ratio did not dramatically exceed the proposed cap, I did not impose this limitation on the optimal allocation models.

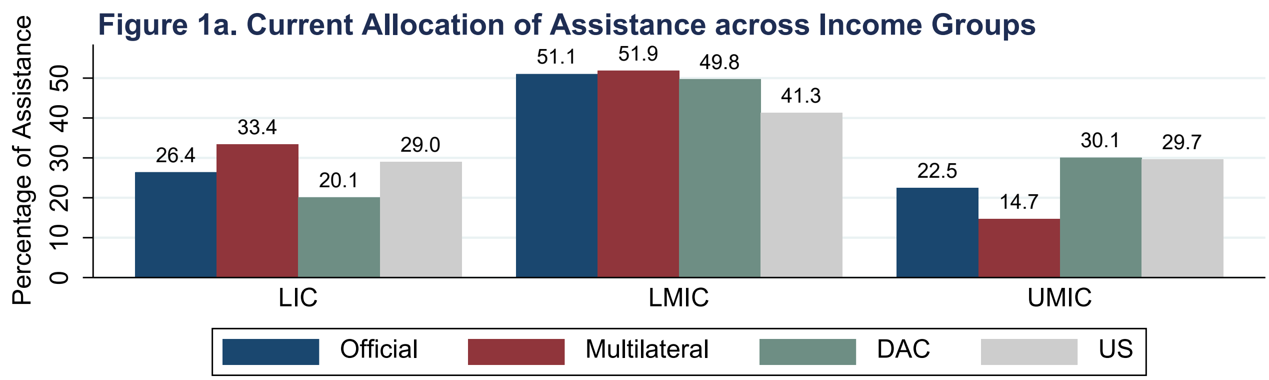
Findings

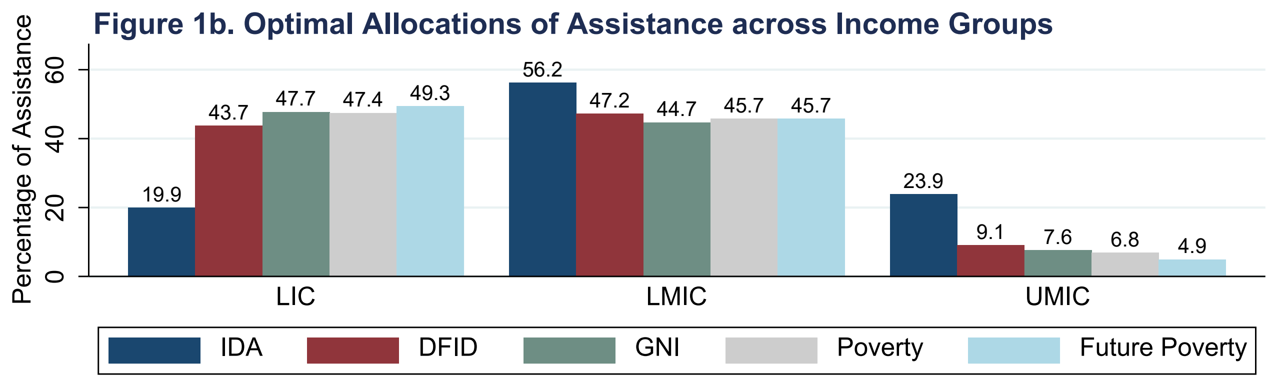
The first section of my findings compares the actual allocation of assistance to the optimal allocation models across income groups. The second section compares the actual allocations to the optimal allocation models across countries. The third section calculates the extent to which countries are under- or over-aided relative to the optimal allocations.

Current versus optimal allocations by income group

Figure 1 shows the distribution of the allocation of assistance across income groups. Figure 1a shows the current allocation for official donors (total aid), multilateral donors, DAC donors, and the United States. In general, most assistance goes to LMICs, and LICs and UMICs receive about equal proportions. Across all groupings of donors, LMICs receive the largest proportion of assistance. The weighted average for all donors is slightly more than half (51.1 percent), which is similar to multilateral donors and the DAC. Only the US gives significantly less than half of its CPA to LMICs. LICs are the group that receives the second most assistance overall, but this is not the case for DAC donors or the US. While LICs receive slightly more than one quarter of CPA overall, LICs receive only about 20 percent of bilateral CPA from the DAC. Both the DAC and the US give relatively more of their CPA to UMICs, at around 30 percent.

[65] Burundi only exceeded the absorptive capacity limit for the GNI and future poverty models and only by roughly one to three percentage points, i.e., total aid was 31 to 33 percent of GDP depending on the model.





Source: WDI, CPR, WGI, PIP, World Poverty Clock.
Note: Actual allocations in top panel; optimal allocations in bottom panale. Aid is CPA averaged over 2019-2021; allocation criteria are averaged 2017-2019.

Figure 1b shows the optimal allocations across income groups for the five models. For most models, the vast majority of assistance is split evenly between LICs and LMICs, except in the IDA model. As mentioned in the last section, the IDA allocation model does not reduce allocations significantly for average income and emphasizes policy performance. Furthermore, countries with relatively lower average income tend to do worse on measures of policy performance. As a result, allocations to LICs are substantially lower for the IDA allocation model than the four others – LICs receive less than UMICs even with the IBRD threshold put in place that mean many UMICs receive no assistance at all. When the DFID model changes the exponent on the need parameter, it greatly changes this distribution. The proportion going to LICs more than doubles from 20 percent to nearly 44 percent and the proportion going to UMICs drops from nearly one quarter (23.9 percent) to less than one tenth (9.1 percent).

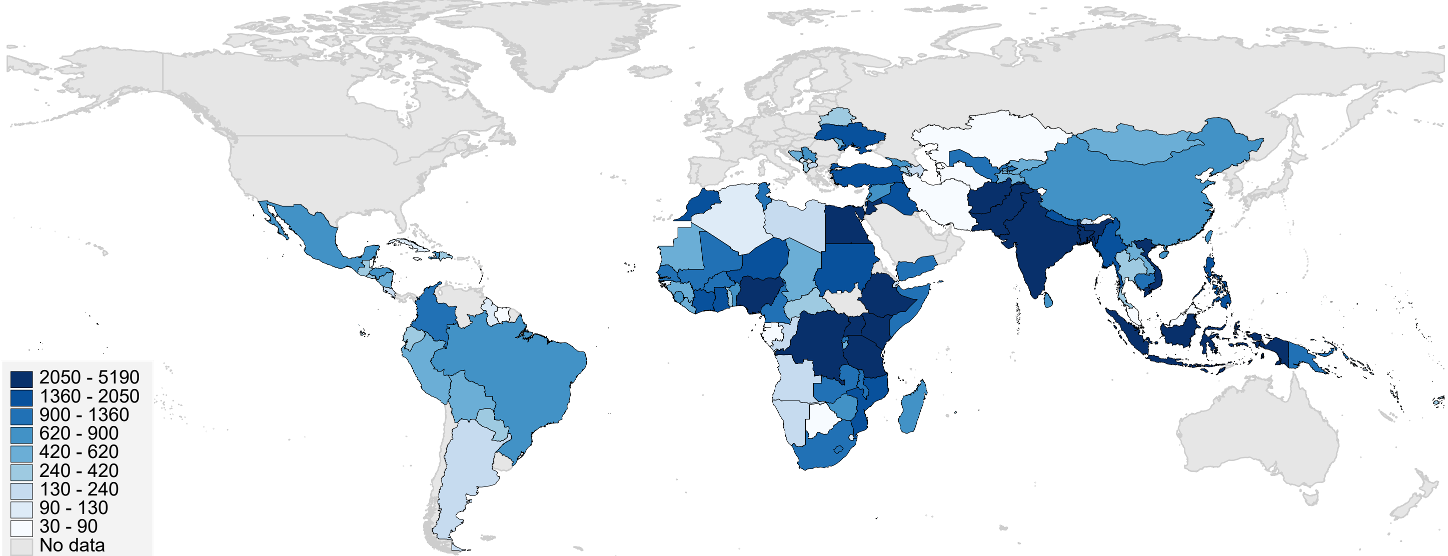
The DFID allocation model is more in line with the needs-based approaches, which all approximately evenly distribute aid between LICs and LMICs. This is partially because of Nigeria and India, which are huge LMICs with significant numbers of people still living in extreme poverty. Despite the allocation models heavily favoring lower average income, LMICs still receive about 45 percent of resources in all of the needs-based approaches, which is only slightly lower than LICs. The differences between the DFID model and the GNI model are notable, as the difference represents the adjustment made by policy performance (as the need parameter is exactly the same and the DFID model only adds the CPR[3] parameter). Essentially the performance-based parameter in the DFID model shifts four percentage points of the global aid share from LICs to LMICs and UMICs.

The allocation model based on a country's share of the global poor is very similar to the GNI model. This is not surprising given the strong correlation between average income and poverty measures shown in the second chapter. As might be expected, global poverty is projected to persist primarily in the poorest countries when we look ahead to 2030. Therefore, the model based on projected poverty slightly shifts the allocation of resources (relative to the GNI and extreme poverty models) by about two percentage points from UMICs to LICs. This is the most progressive allocation of resources from the perspective of income groups, as LICs receive nearly half of assistance – the most of any optimal allocation model – and UMICs receive the least of any model at about five percent.

Current versus optimal allocations by country

Figure 2 maps the distribution of total official assistance, and Table 2 shows the ten countries with the largest allocation shares for both the current allocations of multilateral donors, the DAC, and the US and the five optimal allocation models. This is shown in allocation shares to make the different donor groupings comparable. I also show the income group of each country. The largest recipient of official assistance is Bangladesh, which is the top multilateral recipient and second DAC recipient, though they are not among the top US recipients. Bangladesh is followed closely by India for total aid, which is the top recipient of DAC assistance. Pakistan and Ethiopia are the second and third largest multilateral recipients, and they also do not make the US list – Ethiopia is the tenth largest recipient of DAC assistance, and Pakistan does not make the DAC's top ten either. While there are no UMICs among the top multilateral recipients, Indonesia and Jordan are third and fifth on the DAC list, respectively, and Jordan is the top recipient of US assistance. South Africa is another UMIC that is the third largest US recipient, and Colombia is the tenth largest US recipient. This is consistent with the observation above that multilateral donors give less assistance to UMICs than bilateral donors.

Figure 2. Total Official Aid



Note: Total amount of CPA from all official donors averaged over 2019-2021.
Source: OECD.

I show the same analysis for aid per capita in Appendix 1.[66] The multilateral and DAC lists for aid per capita are dominated by small islands states due to the small-country bias with a few strategic partners

[66] I could conduct the same analysis for total aid amounts as well, but the top recipients in terms of total volumes would be exactly the same as in Table 2.

interspersed for the US. On the other hand, the optimal allocation models reveal their tendencies given that the population parameter in the IDA and DFID model wipe out the per capita aspect of aid per capita. The IDA model favors the best-governed countries overall, and the DFID model favors the best-governed poor countries. The GNI model is an ordered list of the world's poorest countries, the extreme poverty model is a list of the countries with the highest poverty headcount ratios, and the projected poverty model lists the countries with the highest projected poverty rates in the future. While the actual countries on the lists are less important, particularly given the over-representation of small island states for the actual allocations, it is important to keep in mind the tendencies of each model.

Table 2. Top 10 Country Partners in Aid Shares for Actual Donor Allocations and Optimal Allocation Models

Multilateral (Actual)			DAC Donors (Actual)			United States (Actual)			IDA Formula (Model)		
Country	Aid	Income Group	Country	Aid	Income Group	Country	Aid	Income Group	Country	Aid	Income Group
Bangladesh	5.0%	LMIC	India	7.3%	LMIC	Jordan	7.1%	UMIC	Indonesia	10.0%	UMIC
Pakistan	4.7%	LMIC	Bangladesh	4.8%	LMIC	Afghanistan	6.9%	LIC	Pakistan	8.0%	LMIC
Ethiopia	4.4%	LIC	Indonesia	3.7%	UMIC	S. Africa	4.7%	UMIC	Nigeria	5.6%	LMIC
Kenya	4.2%	LMIC	Afghanistan	3.5%	LIC	Kenya	4.5%	LMIC	Ethiopia	5.2%	LIC
Nigeria	3.7%	LMIC	Jordan	3.2%	UMIC	Nigeria	4.0%	LMIC	India	5.0%	LMIC
DRC	2.9%	LIC	Philippines	3.1%	LMIC	Tanzania	4.0%	LMIC	Bangladesh	4.0%	LMIC
Tanzania	2.4%	LMIC	Kenya	2.4%	LMIC	Uganda	3.8%	LIC	Philippines	3.7%	LMIC
Sudan	2.4%	LIC	Vietnam	2.3%	LMIC	Mozambique	3.1%	LIC	Vietnam	3.3%	LMIC
Uganda	2.4%	LIC	Myanmar	2.3%	LMIC	Zambia	2.9%	LMIC	Egypt	2.9%	LMIC
Afghanistan	2.3%	LIC	Ethiopia	2.2%	LIC	Colombia	2.9%	UMIC	Kenya	2.4%	LMIC

DFID Formula (Model)			Average Income (Model)			Extreme Poor (Model)			Future Extreme Poor (Model)		
Country	Aid	Income Group	Country	Aid	Income Group	Country	Aid	Income Group	Country	Aid	Income Group
Ethiopia	9.5%	LIC	DRC	8.8%	LIC	Nigeria	12.2%	LMIC	Nigeria	21.3%	LMIC
Pakistan	8.0%	LMIC	Ethiopia	7.0%	LIC	DRC	10.6%	LIC	DRC	13.7%	LIC
DRC	5.6%	LIC	Pakistan	6.9%	LMIC	India	5.0%	LMIC	Madagascar	4.7%	LIC
India	5.0%	LMIC	India	5.0%	LMIC	Tanzania	5.0%	LMIC	Mozambique	3.6%	LIC
Indonesia	4.5%	UMIC	Nigeria	4.8%	LMIC	Ethiopia	4.3%	LIC	Tanzania	3.6%	LMIC
Nigeria	4.4%	LMIC	Bangladesh	4.1%	LMIC	Madagascar	4.2%	LIC	India	2.9%	LMIC
Uganda	3.7%	LIC	Indonesia	3.5%	UMIC	Mozambique	3.7%	LIC	S. Africa	2.9%	UMIC
Mozambique	3.5%	LIC	Afghanistan	3.5%	LIC	Uganda	3.5%	LIC	Malawi	2.8%	LIC
Bangladesh	3.2%	LMIC	Mozambique	3.1%	LIC	Yemen	3.3%	LIC	Burundi	2.5%	LIC
Tanzania	3.0%	LMIC	Uganda	2.8%	LIC	Indonesia	2.9%	UMIC	Zambia	2.4%	LMIC

Source: WDI, CPR, WGI, PIP, Cuaresma (2018).
Note: Aid is CPA averaged over 2019-2021; allocation criteria are averaged 2017-2019; income groups are from 2019.

The optimal allocation models in Table 2 produce substantially different lists. For instance, the top country from the IDA allocation model is Indonesia, a well-governed UMIC, that does not appear among the top multilateral or US recipient, though it is third for the DAC as a whole. This points to the weakness of the IDA allocation model – it strongly favors better-governed countries while only slowly drawing down assistance at higher average incomes. As a result, well-governed UMICs with large populations tend to do very well, though they are not at the top of the list for most donors.[67]

The DFID allocation model more closely resembles the multilateral donor allocation with two of the same top three countries. It tends to favor large-population poor countries, such as Ethiopia, DRC, and India. The GNI model (without the governance parameter) produces a similar list, but some of the poorly governed countries move up (DRC and Nigeria), while some of the better-governed countries move down (Indonesia and Uganda). Naturally, the extreme poverty model reflects where the greatest number of poor currently live – India would top the list if it were not capped at 5 percent of the allocation. The list is dominated by African countries – all except India, Yemen (ninth), and Indonesia (tenth). The future poverty model has even more of an African bias with the only non-African country being India (sixth). This is in line with projections that the future of global extreme poverty will overwhelmingly be in Africa (World Bank 2020).

Over- and under-aided countries

Finally, I compare current allocations to optimal allocations on a country-by-country basis to see where the largest discrepancies lie. Pietschmann (2014) also examined under-aided countries with three motivations: efficiency gains can be realized by shifting from over-aided to under-aided countries; the

[67] To be fair, the IDA model is only used for IDA-eligible countries (with a graduation process that starts at $1,945 in 2019), which are mostly LICs and relatively poor LMICs. In other words, the allocation model is not designed to apply to UMICs (except the few that remain IDA-eligible), as it typically cuts off assistance well before a country reaches UMIC status.

perception that the poorest countries are over-looked by donors; and under-aided countries might impose negative cross-border spillovers on their neighbors or undermine global public goods. More operationally, OECD (2013) produced a "watch list" of under-aided countries, which consisted primarily of fragile states in Africa. Mitchell and Hughes (2020) set up a test to determine which countries are most "under-aided" and "over-aided" for the purposes of measuring how much individual donors contribute to the mis-allocation problem and help them understand where they could better focus their aid allocations. I am most interested in total official aid and the US allocation, so I present the differences only vis-à-vis those allocations to keep the analysis manageable.

Table 3. Top 10 Under- and Over-Aid Country Partners by Total Volume of Aid for Official Donors

IDA Formula		DFID Formula		Average Income		Extreme Poor		Future Poverty	
Country	**Over**	**Country**	**Over**	**Country**	**Over**	**Country**	**Over**	**Country**	**Over**
Afghanistan	2,046.8	Jordan	2,234.3	Jordan	2,277.3	Egypt	3,203.7	Bangladesh	4,974.0
Jordan	1,995.2	Egypt	1,757.4	Kenya	1,806.1	Jordan	2,411.1	Egypt	3,646.7
Sudan	1,530.4	Bangladesh	1,615.4	Egypt	1,685.5	Bangladesh	2,328.4	Pakistan	2,800.1
Somalia	924.1	Sudan	1,545.4	WB & Gaza	1,324.5	Vietnam	1,979.9	Jordan	2,411.3
Tunisia	890.2	Tunisia	1,137.8	Tunisia	1,183.8	Myanmar	1,697.6	Vietnam	2,146.0
Yemen	874.9	Kenya	1,081.0	Morocco	982.3	Ukraine	1,532.4	India	1,878.5
Kenya	828.8	Iraq	1,080.1	Iraq	934.4	Morocco	1,497.9	Indonesia	1,819.2
Bangladesh	763.1	Morocco	872.9	Colombia	886.9	WB & Gaza	1,379.2	Morocco	1,559.9
China	757.2	PNG	820.5	Cote d'Ivoire	805.4	Iraq	1,364.7	Philippines	1,549.7
Cambodia	753.7	Colombia	774.5	PNG	768.4	Tunisia	1,357.2	Ukraine	1,530.2

IDA Formula		DFID Formula		Average Income		Extreme Poor		Future Poverty	
Country	**Under**	**Country**	**Under**	**Country**	**Under**	**Country**	**Under**	**Country**	**Under**
Indonesia	-8,743.9	Ethiopia	-7,047.6	DRC	-7,579.4	Nigeria	-10,728.8	Nigeria	-20,738.7
Pakistan	-5,638.1	Pakistan	-5,612.7	Pakistan	-4,382.8	DRC	-9,582.4	DRC	-12,961.3
Nigeria	-3,401.1	DRC	-4,090.9	Ethiopia	-4,291.7	Madagascar	-3,700.3	Madagascar	-4,304.9
Ethiopia	-2,335.9	Indonesia	-2,705.0	Nigeria	-2,519.2	Tanzania	-3,193.7	Angola	-2,386.9
Philippines	-2,270.6	Nigeria	-2,052.7	Burundi	-2,136.3	Yemen	-2,458.0	Burundi	-2,345.8
Thailand	-2,151.0	Mozambique	-1,938.0	Madagascar	-2,071.7	Mozambique	-2,163.6	Mozambique	-2,089.6
Iran	-2,093.8	Uganda	-1,837.5	Indonesia	-1,608.7	Angola	-1,829.2	S. Africa	-2,046.5
Vietnam	-1,437.5	Madagascar	-1,623.4	Mozambique	-1,492.1	Uganda	-1,664.6	Malawi	-1,984.2
S. Africa	-1,289.1	Niger	-1,318.5	Malawi	-1,114.7	Malawi	-1,618.9	Zambia	-1,704.0
Algeria	-1,112.2	Malawi	-1,197.0	Uganda	-906.7	S. Africa	-1,520.5	Tanzania	-1,663.1

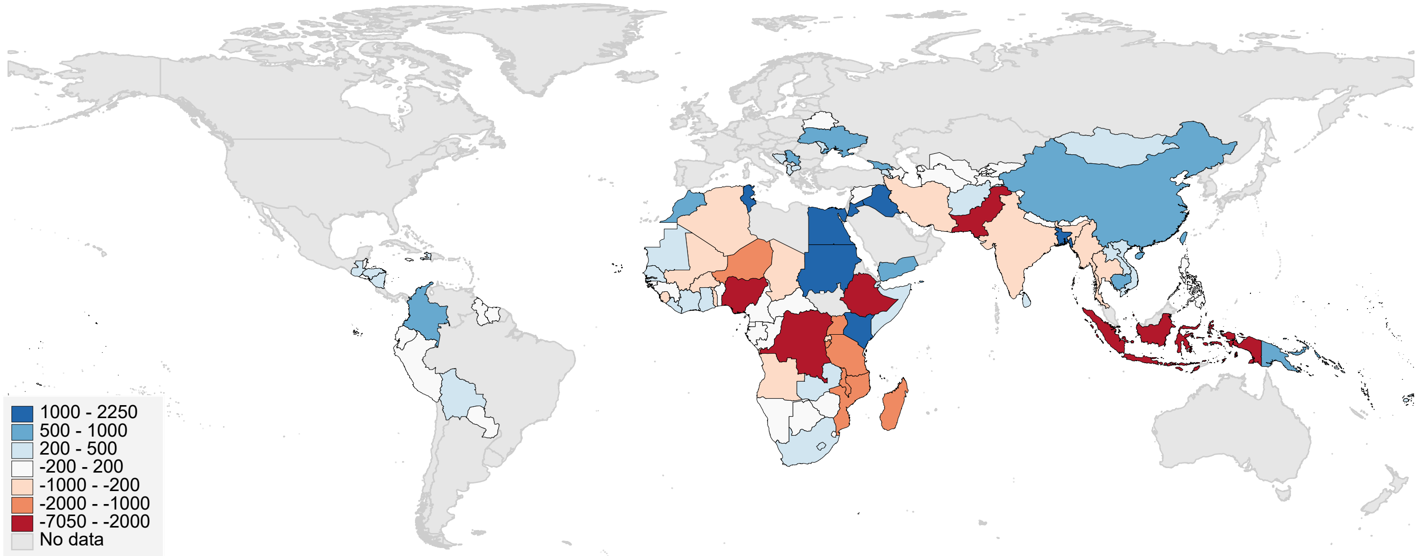
Source: WDI, CPR, WGI, PIP, World Poverty Clock.

Note: Aid is CPA averaged over 2019-2021; allocation criteria are averaged 2017-2019; income groups are from 2019.

Table 3 presents the most over-aided and under-aided countries for total official assistance according to the five different optimal allocation models. The IDA model shows that Afghanistan and Jordan are the most over-aided countries, which is not surprising given their status as geopolitically important partners. The rest of the list is largely populated by other strategic partners and conflict-affected states (e.g., Somalia and Yemen). There is significant overlap between the IDA formula and the other formulas, which are mainly comprised of strategic partners in the Middle East, such as Egypt, Jordan, and Iraq. Jordan appears in the top four of all five models, and Egypt is in the top three for all models except IDA. These lists are mainly driven the by large flows of resources going to these countries as opposed to differences in the underlying models. There is more heterogeneity in the under-aided countries.

The under-aided countries list from the IDA formula is largely dominated by large-population LMICs and UMICs, such as Indonesia and Thailand. The DFID model provides a very different list. Ethiopia is at the top, and eight of the ten countries are African states. DRC and Nigeria are near the top of all three needs-based allocation models, which are also dominated by African countries. Indeed, all ten countries from the future poverty allocation model are in Africa.

Figure 3. Over- and Under-Aided Countries of Total Official Aid



Source: WDI, CPR, WGI, PIP, World Poverty Clock.
Note: Total amount of CPA from all official donors over or under the DFID allocation model. Aid is CPA averaged over 2019-2021; allocation criteria are averaged 2017-2019.

My preferred model is the DFID model, which takes account of governance, but places a stronger emphasis on country need. I map the results of the DFID model for all official aid in Figure 3. Notably, much of sub-Saharan Africa and South Asia is under-aided. The Middle East, Eastern Europe, and Central America all generally receive too much assistance according to the DFID model.

I now focus on US assistance. Table 4 shows the under-aided and over-aided countries for US assistance according to the five optimal allocation models. There is more unanimity across the allocation models for US assistance than all official assistance. Jordan is the most over-aided country for every single allocation model by $840 to $877 million. Afghanistan is in the top three on all lists except current extreme poverty, and South Africa is also in the top five for all five models. Colombia and Iraq also show up for all of the models except for IDA.

The most under-aided countries differ much more by the allocation model than the over-aided countries, and these lists more closely resemble that of official aid in Table 3. Like above, Indonesia and

Pakistan are the most under-aided countries for US assistance according to the IDA model. Also, Ethiopia and Pakistan top the DFID model's list of under-aided countries. The top three countries for the GNI, extreme poverty, and projected poverty models are also almost exactly the same. The DRC and Pakistan are most under-aided in terms of GNI, and DRC and Nigeria are the most under-aided when it comes to the poverty models, both current and future.

Table 4. Top 10 Under- and Over-Aid Country Partners by Total Volume of Aid for the United States

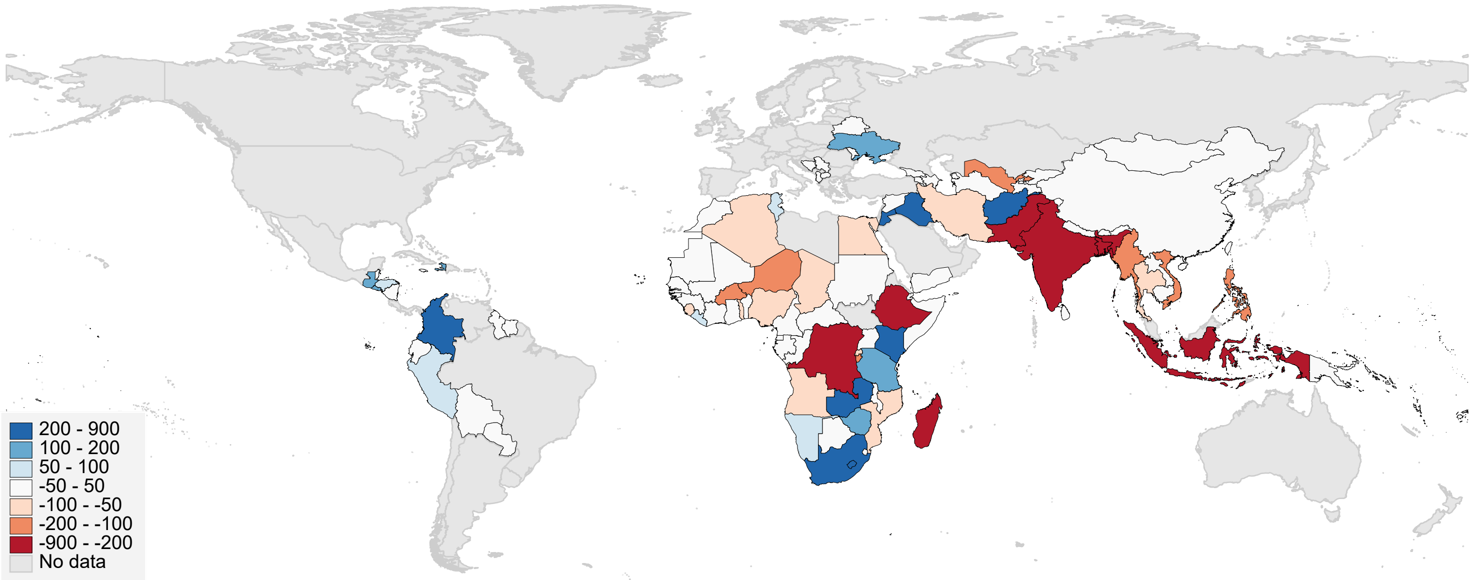
IDA Formula		DFID Formula		Average Income		Extreme Poor		Future Poverty	
Country	**Over**	**Country**	**Over**	**Country**	**Over**	**Country**	**Over**	**Country**	**Over**
Jordan	830.2	Jordan	857.0	Jordan	861.8	Jordan	876.8	Jordan	876.8
Afghanistan	732.2	Afghanistan	555.1	S. Africa	519.7	Colombia	299.7	Afghanistan	574.9
S. Africa	303.3	S. Africa	497.2	Afghanistan	420.8	S. Africa	277.4	Kenya	333.6
Zambia	285.1	Colombia	296.3	Kenya	368.0	Iraq	241.6	Colombia	298.6
Kenya	258.5	Kenya	286.8	Colombia	308.9	Kenya	213.5	Iraq	224.5
Tanzania	232.6	Zambia	276.6	Zambia	278.8	Ukraine	211.5	S. Africa	218.5
Mozambique	225.6	Iraq	209.7	Iraq	193.4	El Salvador	132.9	Ukraine	211.3
Uganda	224.9	Haiti	161.4	Tanzania	146.1	Haiti	118.2	Pakistan	196.5
Haiti	163.5	El Salvador	123.8	Haiti	138.0	Guatemala	110.3	Uganda	174.3
Malawi	161.5	Guatemala	116.5	El Salvador	125.6	Tunisia	104.4	Egypt	151.3

IDA Formula		DFID Formula		Average Income		Extreme Poor		Future Poverty	
Country	**Under**	**Country**	**Under**	**Country**	**Under**	**Country**	**Under**	**Country**	**Under**
Indonesia	-1,104.0	Ethiopia	-862.1	DRC	-835.5	DRC	-1,059.8	Nigeria	-2,140.7
Pakistan	-748.7	Pakistan	-745.9	Pakistan	-608.1	Nigeria	-1,019.4	DRC	-1,438.3
India	-522.0	India	-522.0	Ethiopia	-553.4	India	-522.0	Madagascar	-503.4
Bangladesh	-350.8	DRC	-444.7	India	-522.0	Madagascar	-435.7	Burundi	-275.1
Philippines	-342.4	Indonesia	-427.5	Bangladesh	-362.4	Yemen	-349.0	India	-266.1
Ethiopia	-334.3	Bangladesh	-255.3	Indonesia	-304.7	Uzbekistan	-243.0	Angola	-252.5
Vietnam	-296.3	Madagascar	-203.0	Madagascar	-253.2	Indonesia	-233.7	Chad	-195.1
Thailand	-252.3	Niger	-188.4	Burundi	-251.6	Ethiopia	-217.1	Sudan	-164.7
Iran	-242.5	Burkina Faso	-149.7	Somalia	-163.5	Angola	-190.1	Yemen	-152.3
Nigeria	-198.6	Myanmar	-145.3	Myanmar	-158.7	Somalia	-180.2	Niger	-131.0

Source: WDI, CPR, WGI, PIP, World Poverty Clock**.**
Note: Aid is CPA averaged over 2019-2021; allocation criteria are averaged 2017-2019; income groups are from 2019.

Again, I map the results for the DFID model in Figure 4 – this time for US assistance. While Central Africa and South Asia remain under-aided, almost all of South Asia (except for Afghanistan) are also under-aided. Like total official assistance, Central America and Southern Africa are also largely over-aided for US assistance.

Figure 4. Over- and Under-Aided Countries of United States Aid



Source: WDI, CPR, WGI, PIP, World Poverty Clock.
Note: Total amount of CPA from the US over or under the DFID allocation model. Aid is CPA averaged over 2019-2021; allocation criteria are averaged 2017-2019.

Conclusion and policy implications

My results suggest that aid is substantially mis-allocated according to any of the optimal allocation models, and this confirms previous findings, e.g., Collier and Dollar (2000). I find that about half of total official assistance goes to LMICs, while UMICs receive almost one quarter. The US allocates roughly 30 percent of its assistance to UMICs. This is a subjective judgment, but that is almost certainly far too much, and this is backed up by the optimal allocation models. According to my optimal allocation models, about 90 percent of total assistance should be evenly split across LICs and LMICs. That is, donors are substantially over-invested in UMICs. Kenny (2021) agrees that "aid is currently insufficiently focused on the poorest countries," and the optimal allocation models of Mitchell and Hughes (2020) also found that significantly more aid should be allocated to LICs and away from UMICs.

At a country level, the largest recipient of official assistance is Bangladesh followed closely by India, and Jordan is the top recipient of US assistance. I find that the IDA allocation model is not fit-for-purpose to extend into UMICs, as it favors UMICs much more than the other models given their relatively better performance on the governance criteria. The DFID allocation model more closely resembles the multilateral donor allocation and tends to favor large-population poor countries, such as Ethiopia, DRC, and India. The GNI model produces a similar list of priority countries, but some of the poorly governed countries move up, while some of the better-governed countries move down. Naturally, the extreme poverty model reflects where the greatest number of extreme poor currently reside and is dominated by African countries. The future poverty model has even more of an African bias with the only non-African country being India. This is in line with observations that extreme poverty is increasingly concentrated in Africa and with projections that the future of global poverty will overwhelmingly be in Africa (World Bank 2022). This suggest that a strong focus on Africa is appropriate for aid allocation.

Finally, the allocation models show that the most over-aided countries are primarily strategic partners in the Middle East, such as Egypt, Jordan, and Iraq. The under-aided countries are primarily in Africa with Ethiopia, DRC, and Nigeria among the most under-aided countries. For US assistance, Jordan, Afghanistan, and South Africa are among the most over-aided countries. Colombia and Iraq are also considered highly over-aided for most models. Ethiopia and Pakistan top the DFID model's list of under-aided countries, whereas DRC Nigeria, and Pakistan are the most under-aided according to the needs-based models. Again, these findings strongly suggest an increasing focus on Africa. Particularly for the US, the optimal allocation models a dramatic shift away from strategic partners towards large-population poor countries in Africa.

Most donors would benefit from more explicit allocation criteria (Ottersen 2017). While Mitchell and Hughes (2020) admit that there is no agreed optimal allocation model and the OECD (2013) generously observed that current allocation models are complex, more transparent criteria could help most official donors, nonetheless. Although this is not an exact science, being forced to state a set of allocation principles, if not explicit criteria, is helpful for both transparency and discipline. Transparency is important for accountability as well as the opportunity to deliberate on improvements and learn from experience and new evidence. By stating allocation principles, this can help to limit the amount of bureaucratic infighting and political wrangling that typically characterize an inter-agency budget process, particularly among highly fragmented bilateral donors such as the US. The question then becomes which model is best, and this chapter provides the analysis needed to make an informed decision. In general, the IDA model favors the best-governed countries overall, which tend to be richer countries, and the DFID model favors the best-governed poor countries. The GNI model favors the world's poorest countries, the extreme poverty model favors countries with the highest poverty headcount ratios, and the projected poverty model favors countries with the highest projected poverty rates in the future. Once a policymaker understands the trade-offs and implications of each model, they can make an evidence-based choice of the allocation model that best suits their preferences and constraints.

In addition, donors should establish a more explicit threshold for graduation from assistance. Very few donors implement such a policy, and even the well-known IDA threshold is distorted with exceptions and drawn-out graduation processes. As an example of such a policy, MCC is currently prohibited from working with UMICs – that is, once a country shifts from LMIC to UMIC status, they can no longer be considered to be selected to develop a new program. While this threshold is sometimes criticized as dogmatic or overly strict (Rose *et al* 2016), it is more flexible than realized – a country only needs to be an LMIC in the year of selection, and MCC can continue to develop a program even after a country graduates. This suggests that it is possible to have a hard cut-off in terms of average income without completely disrupting partner country relationships. While there is not a robust literature on donor exits, d'Orey and Prizzon (2019) discuss donor management of the transition away from assistance and eventual exit. The authors examine multiple donors and find that there is no formal approach for most of them – much like aid allocation criteria. More formal rules around graduation may be useful, at least partially because this raises issues related to country need and forces a donor to identify its allocation criteria and make a determination of when they believe aid is no longer needed. This is an extension of policy decisions regarding allocation criteria, and it is a natural consequence of defining those criteria more explicitly, even though it is often overlooked or ignored by donors.

Finally, if a donor a prefers a performance-based approach but also does not want to neglect the neediest countries, a pool of resources could be set aside specifically for very poor or fragile states regardless of policy performance. This suggestion is in response to the objection to performance-based approaches that fragile states are discriminated against because of the effectiveness term even though the future of global poverty is likely to be in fragile states (Milante *et al* 2016). Proponents of aid to fragile states often argue that citizens should not be punished for their corrupt or ineffective governments. Relatedly, Bourguignon and Platteau (2022) propose a "need-adjusted aid effectiveness" approach to allocation that recommends that donors should not consider governance in the very poorest countries. However, this returns to the same effectiveness problem in that the resources may not be spent well. While the authors suggest that such funding should look to improve policy or influence project management, that approach may be too ambitious. While direct service delivery can undermine government effectiveness and institutional development, cash transfers can put resources into the hands of those that need it, stimulate local markets, and are proven to be effective in reducing poverty. Critics of such an approach charge that it does not promote broad development progress, but in the absence of opportunities to effectively promote state-led development, direct transfers of resources to households may be a second-best solution.

Throughout this chapter, I have repeatedly come back to the implications flowing from each of the optimal allocation models. In this way, my analysis helps inform policymakers on which allocation model might make the most sense for their circumstances. That is, the evidence produced in this and previous chapters can and should help to inform policy choices regarding the most appropriate aid allocation model. Once an allocation model is chosen, this type of analysis also provides a list of the highest priority countries for which to increase or decrease assistance. While my analysis and recommendations focus mainly on the US aid system, these insights could also be applied to other donors. Implementing such a data-driven and disciplined approach to allocation policy could pay huge dividends when it comes to maximizing the impact of assistance.

Appendix 1. Top 10 Country Partners by Aid per Capita

The same analysis conducted in Table 2 is repeated below in Table A1 for aid per capita. The optimal allocation models are allocating total official aid. As described above, most of the top recipients of the actual allocation of assistance (for multilateral, DAC, and US) in terms of aid per capita are small island states. On the other hand, the tendencies of the allocation models are laid bare when the population parameter is effectively stripped out.

Table A1. Top 10 Country Partners in Aid per capita for Actual Donor Allocations and Optimal Allocation Models

Multilateral (Actual)			DAC Donors (Actual)			United States (Actual)			IDA Formula (Model)		
Country	Aid per capita	Income Group	Country	Aid per capita	Income Group	Country	Aid per capita	Income Group	Country	Aid per capita	Income Group
Tuvalu	1,568.7	UM	Tuvalu	1,599.7	UM	Marshall Is.	1,313.1	UM	Samoa	78.8	UMIC
Dominica	864.9	UM	Marshall Is.	1,534.5	UM	Micronesia	795.6	LM	Rwanda	77.7	LIC
St Vincent	666.6	UM	Micronesia	896.1	LM	Jordan	88.1	UM	Bhutan	71.4	LMIC
Tonga	649.6	UM	Tonga	512.6	UM	Eswatini	48.8	LM	Cabo Verde	69.0	LMIC
Marshall Is.	491.2	UM	Kiribati	404.5	LM	Lesotho	33.3	LM	Senegal	57.5	LMIC
Grenada	466.0	UM	Samoa	334.5	UM	Liberia	28.8	L	Ghana	57.2	LMIC
Saint Lucia	390.7	UM	Vanuatu	261.9	LM	Namibia	26.6	UM	Benin	56.0	LMIC
Samoa	308.5	UM	Solomon Is.	246.8	LM	Afghanistan	23.1	L	Ethiopia	52.9	LIC
Sao Tome	248.6	LM	Fiji	204.2	UM	Botswana	22.4	UM	Botswana	52.6	UMIC
Maldives	216.8	UM	Mauritius	191.9	UM	El Salvador	21.1	UM	Cote d'Ivoire	51.4	LMIC
DFID Formula (Model)			**Average Income (Model)**			**Extreme Poor (Model)**			**Future Extreme Poor (Model)**		
Country	**Aid per capita**	**Income Group**	**Country**	**Aid per capita**	**Income Group**	**Country**	**Aid per capita**	**Income Group**	**Country**	**Aid per capita**	**Income Group**
Rwanda	142.5	LIC	Burundi	229.2	LIC	Madagascar	175.0	LIC	Burundi	248.0	LIC
Burundi	138.6	LIC	Somalia	139.9	LIC	Burundi	160.7	LIC	CAR	222.5	LIC
Mozambique	131.3	LIC	CAR	125.0	LIC	CAR	150.4	LIC	Madagascar	198.1	LIC
Malawi	126.8	LIC	Malawi	122.2	LIC	Malawi	150.0	LIC	DRC	179.4	LIC
Sierra Leone	126.3	LIC	Mozambique	116.2	LIC	Somalia	149.8	LIC	Malawi	170.2	LIC
Niger	120.3	LIC	DRC	115.4	LIC	DRC	139.2	LIC	Liberia	169.3	LIC
Burkina Faso	98.7	LIC	Madagascar	113.0	LIC	Mozambique	139.0	LIC	Zambia	153.9	LMIC
Ethiopia	96.1	LIC	Sierra Leone	110.8	LIC	Zambia	134.2	LMIC	Guinea-Bissau	146.8	LIC
Madagascar	96.0	LIC	Afghanistan	103.8	LIC	Yemen	127.1	LIC	Mozambique	136.5	LIC
Uganda	95.1	LIC	Niger	98.8	LIC	Niger	114.2	LIC	Lesotho	123.0	LMIC

Source: WDI, CPR, WGI, PIP, World Poverty Clock.
Note: Actual aid per capita is CPA averaged over 2019-2021; allocation criteria averaged 2017-2019; income groups from 2019.

References

Acemoglu, Daron, Suresh Naidu, Pascual Restrepo, and James A. Robinson. 2019. "Democracy does cause growth." *Journal of political economy* 127, no. 1: 47-100.

Alatas, Vivi, Abhijit Banerjee, Rema Hanna, Benjamin A. Olken, and Julia Tobias. 2012. "Targeting the Poor: Evidence from a Field Experiment in Indonesia." *American Economic Review*, 102 (4): 1206-40.

Alesina, Alberto, and David Dollar. 2000. "Who gives foreign aid to whom and why?." Journal of economic growth 5: 33-63.

Alonso, José Antonio. *Cooperation with middle-income countries: An incentive-based approach*. AECID, 2014.

Anand, Sudhir. 1977. "Aspects of poverty in Malaysia." *Review of Income and Wealth* 23, no. 1: 1-16.

Anderson, Edward. 2008. "Practices and implications of aid allocation." Background Study for the 2008 Development Cooperation Forum. UN Economic and Social Council (ECOSOC).

Andersen, Jørgen Juel, Niels Johannesen, and Bob Rijkers. 2022. "Elite capture of foreign aid: Evidence from offshore bank accounts." *Journal of Political Economy* 130, no. 2: 388-425.

Angrist, Joshua D., and Jörn-Steffen Pischke. 2009. *Mostly harmless econometrics: An empiricist's companion.* Princeton university press.

Arkedis, Jean. 2011. “Getting to a “Grand Bargain” for Aid Reform: The Basic Framework for U.S. Foreign Assistance.” Center for Global Development Essay.

Attridge, Samantha, and Lars Engen. 2019. *Blended finance in the poorest countries: the need for a better approach*. ODI Report.

Azizi, SeyedSoroosh. 2021. "The impacts of workers’ remittances on poverty and inequality in developing countries." *Empirical Economics* 60, no. 2: 969-991.

Barder, Owen Matthew. 2009. "What is poverty reduction?." *Center for Global Development Working Paper* 170.

Benfield, Andrew and Nevila Como. 2019. *Effectiveness to Impact: Study on the application of the effectiveness principles*. European Union Commission Report.

Berthélemy, J. 2006. "Aid allocation: Comparing donor's behaviours." Swedish economic policy review 13, no. 2: 75.

Beynon, Jonathan. 2003. *Poverty efficient aid allocations: Collier/Dollar revisited.* London: Overseas Development Institute.

Bigsten, Arne Lennart, Jean Philippe Platteau, and Sven Tengstam. 2011. "The Aid Effectiveness Agenda: The benefits of going ahead Final Report." *Contract* 2010/256898.

Birdsall, Nancy, and Christian J. Meyer. 2015. "The median is the message: a good enough measure of material wellbeing and shared development progress." Global Policy 6, no. 4: 343-357.

Bold, Tessa, Mwangi Kimenyi, Gemano Mwabu, Alice Ng'ang'a and Justin Sandefu. 2018. "Experimental evidence on scaling up education reforms in Kenya," *Journal of Public Economics*: 1-20.

Boone, Peter. 1996. 'Politics and the effectiveness of foreign aid', *European Economic Review*, vol. 40(2), pp. 289-329.

Bourguignon, François, and Jan Willem Gunning. 2020. "Foreign aid and governance." *The Handbook of Economic Development and Institutions*: 308.

Bourguignon, François, and Jean-Philippe Platteau. 2015. "The hard challenge of aid coordination." *World Development* 69: 86-97.

Brainard, Lael. 2003. "The millennium challenge account and foreign assistance." *The Brookings Review* 21, no. 2: 41.

Briggs, Ryan C. 2020. "Results from single-donor analyses of project aid success seem to generalize pretty well across donors." *The Review of International Organizations* 15, no. 4: 947-963.

Bulman, David, Walter Kolkma, and Aart Kraay. 2015. "Good Countries or Good Projects? : Comparing Macro and Micro Correlates of World Bank and Asian Development Bank Project Performance." Policy Research Working Paper; No. 7245. World Bank.

Burnside, Craig, and David Dollar. 2000. "Aid, Policies, and Growth." *American Economic Review* 90 (4): 847–68.

Carter, Patrick. 2014. "Aid allocation rules." *European Economic Review* 71: 132-151.

Ceriani, Lidia, and Paolo Verme. 2014. "The income lever and the allocation of aid." *The Journal of Development Studies* 50, no. 11: 1510-1522.

Chancel, Lucas, Thomas Piketty, Emmanuel Saez, and Gabriel Zucman, eds. 2022. *World inequality report 2022*. Harvard University Press.

Claessens, Stijn, Danny Cassimon, and Bjorn Van Campenhout. 2009. "Evidence on changes in aid allocation criteria." *The World Bank Economic Review* 23, no. 2: 185-208.

Clemens, Michael A., Steven Radelet, Rikhil R. Bhavnani, and Samuel Bazzi. 2012. "Counting Chickens When They Hatch: Timing and the Effects of Aid on Growth." *American Economic Journal* 122 (561): 590–617.

Collier, P. and David Dollar. 2002. "Aid Allocation and Poverty Reduction." *European Economic Review.* Vol.46 (8), pp. 1475-1500.

Cordella, Tito. 2018. "Optimizing finance for development." World Bank Policy Research Working Paper 8320.

Crespo Cuaresma, Jesus. 2018. "Replication Data for Will the Sustainable Development Goals be fulfilled? Assessing present and future global poverty.", https://doi.org/10.7910/DVN/XS5KQA.

Crespo Cuaresma, Jesús, Wolfgang Fengler, Homi Kharas, Karim Bekhtiar, Michael Brottrager, and Martin Hofer. 2018. "Will the Sustainable Development Goals be fulfilled? Assessing present and future global poverty." *Palgrave Communications* 4, no. 1.

Crosswell, Michael. 1980. "Need as a Criterion for Allocating Development Assistance" *USAID working Paper*.

Crosswell, Michael 2015. "Country Selectivity and Extreme Poverty." *USAID Economics Brief*, USAID, Washington, DC.

Cogneau, Denis, and Jean-David Naudet. 2007. "Who deserves aid? Equality of opportunity, international aid, and poverty reduction." *World Development* 35, no. 1: 104-120.

Crespo Cuaresma, Jesús, Wolfgang Fengler, Homi Kharas, Karim Bekhtiar, Michael Brottrager, and Martin Hofer. 2018. "Will the Sustainable Development Goals be fulfilled? Assessing present and future global poverty." Palgrave Communications 4, no. 1.

d'Orey, Jalles, Maria Ana, and Annalisa Prizzon. *Exit from aid: An analysis of donor experiences*. ODI Report, 2019.

Das, A, J Friedman, Eshani Kandpal E. 2018. "Does involvement of local NGOs enhance public service delivery? Cautionary evidence from a malaria-prevention program in India." *Health Economics*:172-188.

Davies, R. and J. Pickering. 2015. "Making Development Co-operation Fit for the Future: A Survey of Partner Countries", OECD Development Co-operation Working Papers, No. 20.

Denizer, C., Daniel Kaufmann, and Aart Kraay. 2013. "Good countries or good projects? Macro and micro correlates of World Bank project performance." *Journal of Development Economic*105, 288–302.

Lea, Nick, and Stefan Dercon. 2016. "Benchmarking aid allocation: Is the global and multilateral aid allocation poverty-focused." DFID Chief Economist's Office.

Development Initiatives. 2013. *Investments to End Poverty: Real Money, Real Choices, Real Lives.* Development Initiatives Limited.

Dissanayake, Ranil, Charles Kenny, and Mark Plant. 2020. "What is the role of aid in middle-income countries?." CGD Policy Paper 201.

Dollar, David, Tatjana Kleineberg, and Aart Kraay. 2016. "Growth still is good for the poor." *European Economic Review 81*: 68-85.

Dollar, David, and Victoria Levin. 2006. "The increasing selectivity of foreign aid, 1984–2003." *World development* 34, no. 12: 2034-2046.

Dollar, David, and Lant Pritchett. 1998. "Assessing aid-what works, what doesn't, and why."

Dudley, Leonard, and Claude Montmarquette. 1976. "A model of the supply of bilateral foreign aid." *The American Economic Review* 66, no. 1: 132-142.

Dunning, Casey, Jonathan Karver, and Charles Kenny. 2014. "Hating on the Hurdle: Reforming the Millennium Challenge Corporation's Approach to Corruption." Center for Global Development.

Dunning, Casey, Sarah Rose, and Claire McGillem. 2017. "Implementing Ownership at USAID and MCC: A US Agency-Level Perspective." CGD Policy Paper 99. Center for Global Development.

Easterly, William. 2007. "Are aid agencies improving?." *Economic policy* 22, no. 52: 634-678.

Easterly, William. 2011. "Benevolent autocrats." *Work. Pap., Dev. Res. Inst.*, New York Univ., New York.

Easterly, William, Ross Levine, and David Roodman. 2004. "Aid, Policies, and Growth: Comment." *American Economic Review* 94 (3): 774–80.

Easterly, William, and Tobias Pfutze. 2008. "Where does the money go? Best and worst practices in foreign aid." *Journal of Economic Perspectives* 22, no. 2: 29-52.

Edward, Peter and Andy Sumner. 2013. Estimating the Scale and Geography of Global Poverty Now and in the Future: How Much Difference Do Method and Assumptions Make? *World Development*, vol. 58, pp. 67–82.

Feeny, Simon, and Ashton De Silva. 2012. "Measuring absorptive capacity constraints to foreign aid." Economic Modelling 29, no. 3: 725-733.

Fenochietto, Mr Ricardo, and Ms Carola Pessino. 2013. *Understanding countries' tax effort.* International Monetary Fund.

Galiani, Sebastian, Stephen Knack, Lixin C. Xu, and Ben Zou. 2017. The Effect of Aid on Growth: Evidence from a quasi-Experiment. *Journal of Economic Growth* 22 (1): 1–33.

Grover, Jacob. 2009. *US aid: Helping others or helping themselves?* Georgetown University.

Grover, Jacob. 2023. "Motivations for the allocation of Chinese aid." *China Studies Review*.

Guillaumont, Patrick. 2008. "An economic vulnerability index: its design and use for international development policy."

Guillaumont, Patrick. 2011. "Adapting aid allocation criteria to development goals." halshs-00556806.

Gunning, Jan Willem. 2000. "The reform of aid: conditionality, selectivity and ownership*." In Conference on Aid and Development*, pp. 14-23. Swedish International Development Cooperation Agency.

Hansen, Henrik and Finn Tarp. 2001. "Aid and Growth Regressions." *Journal of Development Economic*s 64(2), 547-570.

Hausmann, Ricardo, Dani Rodrik, and Andrés Velasco. 2008. "Growth diagnostics." *The Washington consensus reconsidered: Towards a new global governance*: 324-355.

Hayes-Birchler, A. and Staats, S.J. 2014. "Country selectivity. Principles into Practice." Washington DC: Millennium Challenge Corporation.

Higgins, Sean, and Nora Lustig. 2016. "Can a poverty-reducing and progressive tax and transfer system hurt the poor?." *Journal of Development Economics* 122: 63-75.

Honig, Daniel. 2018. *Navigation by Judgment: Why and when Top Down Management of Foreign Aid Doesn't Work.* Oxford University Press.

Hughes, Sam, and Ian Mitchell. 2023. "How to Reverse the Decline in the Poverty Focus of UK Aid." CGD Note. Washington, DC: Center for Global Development. https://www.cgdev.org/sites/default/files/how-reverse-decline-poverty-focus-uk-aid.pdf.

Imam, Patrick A., and Jonathan RW Temple. 2023. Political Institutions and Output Collapses. No. 2023/036. International Monetary Fund.

Jerven, Morten. 2013. *Poor numbers: how we are misled by African development statistics and what to do about it.* Cornell University Press.

Jolliffe, Dean. 2014. *A measured approach to ending poverty and boosting shared prosperity: concepts, data, and the twin goals*. World Bank Publications.

Jia, S., and Claudia Williamson. 2019. "Aid, policies, and growth: Why so much confusion?" *Contemporary Economic Policy*, 37, 577–599.

Johnson, J. H., and Wasty, S. S. 1993. "Borrower ownership of adjustment programs and the political economy of reform". Vol. 199. World Bank Publications.

Kappel, Rolf. 2017. *Bilateral cooperation of SDC with Middle Income Countries (MICs): proposal for a method to evaluate priorities for aid allocation: Study on behalf of the Swiss Agency for Development Cooperation (SDC), South Cooperation*. ETH Zurich.

Kanbur, Ravi, and Diganta Mukherjee. 2007. "Poverty, relative to the ability to eradicate it: An index of poverty reduction failure." *Economics Letters* 97, no. 1: 52-57.

Kaufmann, Daniel, Aart Kraay, and Pablo Zoido-Lobaton. 2000. "Governance matters." *Finance and Development* 37, no. 2: 10.

Kenny, Charles. 2011. *Getting Better: why global development is succeeding--and how we can improve the world even more*. Basic Books (AZ).

Kenny, Charles. 2014. "The strange and curious grip of country income status on otherwise smart and decent people." CGD Commentary and Analysis. Washington, DC: Center for Global Development. https://www. cgdev. org/blog/strange-andcurious-grip-country-income-status-otherwise-smart-and-decent-people.

Kenny, Charles. 2021. "We Should Be Spending More of Available Aid in Poorer Countries, Not Less." *CGD Working Paper* 564. Washington, DC: Center for Global Development.

Killick, T. 1998. *Aid and the Political Economy of Policy Change* (1st ed.). Routledge.

Le, Tuan Minh, Blanca Moreno-Dodson, and Nihal Bayraktar. 2012. "Tax capacity and tax effort: Extended cross-country analysis from 1994 to 2009." *World Bank Policy Research Working Paper* 6252 .

Manuel, Marcus, Harsh Desai, Emma Samman, and Martin Evans. 2018. *Financing the end of extreme poverty.* ODI Report.

Martinez, Luis R. 2022. "How much should we trust the dictator's GDP growth estimates?." Journal of Political Economy 130, no. 10: 2731-2769.

McGillivray, Mark. 2003. " Descriptive and prescriptive analyses of aid allocation: approaches, issues and consequences." *WIDER Discussion Paper*, No. 2003/21, The United Nations University World Institute for Development Economics Research (UNU-WIDER), Helsinki.

McGillivray, Mark. 2004. "Modelling Aid Allocation: Issues, Approaches, and Results." *World Institute for Development*.

McGillivray, Mark. 2006. "Aid allocation and fragile states." *Fragile States: Causes, Costs and Responses*: 166-184.

White, Howard, and Mark McGillivray. 1995. "How well is aid allocated? Descriptive measures of aid allocation: a survey of methodology and results." Development and Change 26, no. 1: 163-183.

McKee, Caitlin, Catherine Blampied, Ian Mitchell, and Andrew Rogerson. 2020. "Revisiting Aid Effectiveness: A New Framework and Set of Measures for Assessing Aid "Quality". Working Papers 524, Center for Global Development.

McKinlay, Robert D., and Richard Little. 1979. "The US aid relationship: a test of the recipient need and the donor interest models." *Political Studies* 27, no. 2: 236-250.

Milante, Gary, Barry Hughes, and Alison Burt.2016. "Poverty eradication in fragile places: prospects for harvesting the highest hanging fruit by 2030." *Stability: International Journal of Security and Development* 5, no. 1.

Millennium Challenge Act. 2003.

Millennium Challenge Corporation. 2023. *Congressional Budget Justification, FY 2024*. https://www.mcc.gov/resources/doc/report-congressional-budget-justification-fy-2024.

Mitchell, Ian, and Sam Hughes. 2020. "Which Countries Miss Out in Global Aid Allocation." *CGD Notes*, Washington, DC: Center for Global Development.

Mosley, Paul. 1986. "Aid-effectiveness: The micro-macro paradox." Ids Bulletin 17, no. 2: 22-27.

Mosley, Paul, John Hudson, and Arjan Verschoor. 2004. "Aid, poverty reduction and the 'new conditionality'." *The economic journal* 114, no. 496: F217-F243.

Natsios, A. 2010. "The clash of the counter-bureaucracy and development." Washington, DC: Center for Global Development.

Neumayer, Eric. 2003. "What factors determine the allocation of aid by Arab countries and multilateral agencies?." *Journal of Development Studies* 39, no. 4: 134-147.

OECD. 2003. "Rome declaration on harmonization." In High-Level Forum I on aid effectiveness, 24–25 February 2003. OECD, Rome.

OECD. 2005. "Paris declaration on aid effectiveness." In High Level Forum II on aid effectiveness, 28 February–2 March 2005. OECD, Paris.

OECD. 2008. "Accra Agenda for Action." In High Level Forum III on aid effectiveness. OECD, Paris.

OECD. 2011. "Evaluation of the implementation of the Paris Declaration" OECD, Paris.

OECD. 2013. "Identification and monitoring of potentially under-aided countries."

OECD. 2023. OCED.stat.

Organisation of Economic Co-operation and Development. "OECD Development Co-operation Peer Reviews: United States 2016." OECD Development Co-operation Peer Reviews, OECD Publishing, Paris (2016). https://doi.org/10.1787/9789264266971-en.

Ottersen, Trygve, Suerie Moon, and John-Arne Røttingen. 2017. "Distributing development assistance for health: simulating the implications of 11 criteria." *Health Economics, Policy and Law* 12, no. 2: 245-263.

Parks, Bradley C., and Zachary J. Rice. 2013. "Does the "MCC Effect" Exist? Results from the 2012 MCA Stakeholder Survey." *MCA Monitor*.

Paulo, Sebastian, Heiner Janus, and Sarah Holzapfel. 2017. "Thematic aid allocation: what are the benefits and risks?." *German Development Institute, Briefing Paper* 19.

Pritchett, Lant. 2006. "Who is not poor? Dreaming of a world truly free of poverty." *The World Bank Research Observer* 21, no. 1: 1-23.

Pritchett, Lant, and Addison Lewis. 2022. "Economic growth is enough and only economic growth is enough."

Qian, Nancy. 2015. "Making Progress on Foreign Aid." *Annual Review of Economics* 7(1).

Quibria, MG. 2014. "Aid effectiveness: Research, policy and unresolved issues." *Development Studies Research,* 1(1), 75–87.

Quiggin, J., and Mahadevan, R. 2010. "The poverty burden: A measure of the difficulty of ending extreme poverty." *Risk and Sustainable Management Group Working Papers* 151524, University of Queensland, School of Economics, St. Lucia, Queensland, Australia.

Rabinowitz, Gideon. 2015. "Literature Review on aid ownership and participation." Overseas Development Institute report.

Rajan, Raghuram G., and Arvind Subramanian. 2008. "Aid and Growth: What Does the Cross-Country Evidence Really Show?" *Review of Economics and Statistics* 90 (4): 643–65.

Ravallion, Martin. 2010. "Do poorer countries have less capacity for redistribution?." Journal of Globalization and Development 1, no. 2.

Ravallion, Martin. 2011. "On multidimensional indices of poverty." The Journal of Economic Inequality 9: 235-248.

Roodman, David. 2007. "The Anarchy of Numbers: Aid, Development, and Cross-Country Empirics." *World Bank Economic Review* 21(2), 255-277.

Rose, Sarah, Nancy Birdsall, and Anna Diofasi. 2016. "Creating a better candidate pool for the Millennium Challenge Corporation." *CGD Policy Paper* 82.

Savedoff, William D. 2019. *"What Is" country Ownership"?: A Formal Exploration of the Aid Relationship."* No. 519. Center for Global Development, 2019.

Sen, Amartya. 1982. *Poverty and famines: an essay on entitlement and deprivation*. Oxford university press.

Temple, J.R.W. 2010. 'Aid and conditionality' in (D. Rodrik and M. Rosenzweig, eds.), Handbook of Development Economics , Vol. 5, pp. 4415-523, Amsterdam: Elsevier.

USAID. 2015. "Vision for Ending Extreme Poverty."

Usmani, Faraz, Marc Jeuland, Subhrendu Pattanayak. 2018. "NGOs and the effectiveness of interventions," WIDER Working Paper Series wp-2018-59, World Institute for Development Economic Research (UNU-WIDER).

Weiler, Florian, and Franklins A. Sanubi. 2019. "Development and climate aid to Africa: comparing aid allocation models for different aid flows." *Africa spectru*m 54, no. 3: 244-267.

White, Howard, Radhika Menon, and Hugh Waddington. 2018. “Community-driven development: does it build social cohesion or infrastructure? A mixed-method evidence synthesis”. *3ie Working Paper* 30.

Wood, Adrian. 2008. "Looking ahead optimally in allocating aid." *World Development* 36, no. 7: 1135-1151.

World Bank. 2022. *Poverty and Shared Prosperity 2022: Correcting Course.* The World Bank.