The Elusive Quest for Additionality

Patrick Carter, Nicolas Van de Sijpe, and Raphael Calel

Abstract

Development Finance Institutions (DFIs) are frequently asked to demonstrate their additionality—meaning that they make investments that the private sector would not—but what evidence of additionality would look like is rarely articulated. This paper examines potential quantitative and qualitative evidence. We investigate whether it is possible to infer additionality from observational investment data, and show how the demand-led nature of DFIs' business model can create bias in standard statistical techniques used to identify causal effects. Having established that rigorous evidence of additionality may continue to elude us, we discuss circumstantial evidence that would increase confidence that additionality is present, and propose a probabilistic approach to additionality.

Center for Global Development
www.cgdev.org

Working Paper 495
September 2018

The Elusive Quest for Additionality

Patrick Carter
CDC Group

Nicolas Van de Sijpe
Department of Economics, University of Sheffield

Raphael Calel
McCourt School of Public Policy, Georgetown University

This paper was written whilst Paddy Carter was a Research Fellow at the Center for Global Development and does not represent the views of CDC Group. Thanks to Colin Buckley, Alex MacGillivray, Vijaya Ramachandran, Jon Temple, Frank Windmeijer, Adrian Wood, and an anonymous referee for helpful comments and advice.

Corresponding author pcarter@cdcgroup.com.

The Center for Global Development is grateful for contributions from the Bill & Melinda Gates Foundation in support of this work.

Patrick Carter, Nicolas Van de Sijpe, and Raphael Calel, 2018. "The Elusive Quest for Additionality." CGD Working Paper 495. Washington, DC: Center for Global Development. https://www.cgdev.org/publication/elusive-quest-additionality.

The data used in this paper is available here: https://www.cgdev.org/sites/default/files/elusive-quest-additionality-replication-do-files.zip. More information on CGD's research data and code disclosure policy can be found here: www.cgdev.org/page/research-data-and-code-disclosure.

Center for Global Development
2055 L Street NW
Washington, DC 20036

202.416.4000
(f) 202.416.4050

www.cgdev.org

1 Introduction

Hopes of achieving the Sustainable Development Goals rest, in large part, upon diverting some of the trillions of wealth held in rich countries towards investments in capital-scarce, poor countries.

The attractiveness of investment in poor countries has to do with factors such as natural resource endowments, the political and legal environment, the size of local markets and the availability of complementary inputs, including infrastructure and skilled labour. Much of this is hard for foreign development agencies to affect, at least in the short run.

One thing that donor governments and their agents, such as multilateral development banks, can do in the short run is commit their own capital to investments in private enterprises in developing countries. The entities that do this are known as 'development finance institutions' (DFIs), which includes the private-sector windows of development banks and also various bilateral institutions, some of which are investment funds, not banks.[1] DFIs' objectives include having a positive impact on development with their own investments; 'crowding-in' private sector finance into their investments directly as co-investors; and encouraging further investment indirectly, by creating markets and demonstrating to sceptical private investors where good returns can be made in developing countries.

World governments and their DFIs have coalesced around a 'Financing for Development' strategy of using their funds to 'leverage' a far greater sum of private investment, thereby getting "from billions to trillions."[2] But do investments made by DFIs add to the sum of total investment in developing countries, or do DFIs merely displace private investors who would have invested in those projects anyway? There are few more important questions in development policy than this. The success of the 'billions to trillions' strategy rests on whether DFIs are making a substantial difference to the quantity of investment in the developing world, or merely substituting public finance for private. Indirect demonstra-

[1] The page 'Development finance institutions and private sector development' on the OECD website provides a good overview.

[2] The most recent inter-governmental agreement was signed at the Third International Conference on Financing for Development in Addis Ababa in 2015 (the 'Addis Ababa Action Agenda'). The 'billions to trillions' strategy was first articulated in "From Billions to Trillions: Transforming Development Finance Post-2015 Financing for Development: Multilateral Development Finance" prepared jointly by the African Development Bank, the Asian Development Bank, the European Bank for Reconstruction and Development, the European Investment Bank, the Inter-American Development Bank, the International Monetary Fund, and the World Bank Group for the April 18, 2015 Development Committee meeting.

tion effects are illusory if the private sector would have demonstrated where good returns can be made by itself.

Within the field of development finance, making an investment happen that would not have happened otherwise is referred to as 'additionality'. DFIs are routinely asked to demonstrate or measure their additionality, and are routinely criticised for being unable to do so.[3] What acceptable evidence would look like, or how it could be obtained, is rarely articulated. That is the subject of this paper. The primary focus will be on the question of whether quantitative evidence of additionality could be found by applying statistical techniques to investment data, but qualitative evidence will also be analysed.

Our analysis suggests that definitive evidence of additionality may always be elusive, so the paper concludes with a discussion of evidence that might not demonstrate additionality, but which might cause a reasonable person to believe it is present. We propose that DFIs should take a similarly probabilistic approach to evaluating additionality when making investment decisions.

The data that would be required to implement the quantitative methods we explore in this paper are not always available. Most DFIs report their investments to the OECD Creditor Reporting System, but concerns over commercial confidentiality mean that, to preserve anonymity, it is not possible to extract the data matching individual DFI flows to recipient countries. The quantity of investment is also sometimes omitted. Some researchers have privately compiled partial datasets from sources such as DFI annual reports, see e.g. Massa et al. (2016) and Kenny et al. (2018). The quality of the data on other variables that would be needed, such as private sector fixed capital formation, can be questionable in some of the developing countries where DFIs are active.

However, the greater emphasis on DFIs in global development cooperation has led to increased calls for evidence of their additionality, so in anticipation of donors spending money on data gathering exercises, we think it is worth asking what data could tell us.[4]

[3] See for instance: 'Leveraging Aid: a literature review on the additionality of using ODA to leverage private investments' published by the UK Aid Network; 'Six Criteria for Donor Engagement with the Private Sector' published by Countdown 2030; and 'Financing for development post-2015: Improving the contribution of private finance' a study requested by the European Parliament's Committee on Development.

[4] In 2017 the UK Department for International Development issued a tender for researchers to 'design and implement a large-scale, long-term study analysing the impact of DFIs, including CDC, on private sector investment activity', in response to long-standing calls from the National Audit Office for a 'longitudinal study'.

Identifying causal effects from observational data is a fundamental problem of econometrics.[5] The approach we take in this paper is to test standard techniques on simulated datasets in which we know the true extent of additionality, to see how well they recover the truth. One contribution of this paper is to articulate how DFIs operate, and then to translate this understanding into a formal data generating process (DGP), something that we believe has not been done before and which we hope will help sharpen the debate on this topic.

We show that coefficients estimated by OLS and fixed-effects in cross-country investment regressions are a misleading measure of additionality, a result that will not greatly surprise econometricians. The sign and size of the bias depends on the specific selection mechanism DFIs use to select projects to invest in. A more novel result is that some common techniques for distinguishing between correlation and causation in cross-country data, in particular 'system GMM' estimation of a dynamic panel data model and instrumental variable estimation that relies on a supply-push instrument, may also produce misleading results in this context. Whilst the approach we take shows that biases arise under some plausible configurations of our DGP, we do not know which version of our DGP best corresponds to reality, nor how well, so we caution that the magnitudes of the biases we report should not be taken as assertions about the magnitudes of biases when these methods are used in practice.

We also consider the possibility of identifying additionality in firm-level data. We explore the idea of using observable project characteristics to estimate the probability the private investors would have undertaken a project, and we show that such an approach can be misleading for two reasons: investors base their decisions on considerations that cannot be captured by simple descriptive data, which can cause DFIs' and private investments to superficially resemble each other even when DFIs are additional; and DFIs' and private investments may look different if DFIs take all the deals of a certain type, even if they are crowding-out private investors.

Our analysis of the qualitative evidence is similarly pessimistic. We argue self-reported evidence, both from DFIs and their investees, cannot be taken as definitive evidence. We conclude that rather than seeking to 'measure' additionality, we should be asking ourselves

[5] Athey and Imbens (2017) provide an excellent survey of the state of the art.

under what circumstances should we be more or less likely to believe an investment as additional. An approach known as 'process tracing' has the potential to translate qualitative evidence into a probability that additionality is present.[6]

1.1 The Impossibility of Experimentation

Before proceeding to the main analysis, we first briefly discuss why, in this context, quantitative evidence would necessarily come from observational data. Empirical (development) economics has experienced a 'credibility revolution' (Angrist and Pischke, 2010), most notably through the increasing use of randomised control trials to generate reliable estimates of causal effects.[7] In principle, randomising at the deal level, amongst a pool of prospects that DFIs have deemed to be investable and additional, could generate a measure of additionality by revealing how many of the prospects randomly denied investment went on to find finance elsewhere. Randomising DFI investment across countries would capture the full effect of DFI investment on the recipient country, including any possible general equilibrium effects. Unfortunately, we are unlikely ever to have experimental evidence that DFI investments are additional.

The average deal size reported in sub-Saharan Africa in 2015, by the Association of Bilateral European Development Finance Institutions (EDFI), was just under $10m.[8] It is unlikely that anyone would run a trial with amounts of that size invested at random.[9] Besides the sums involved, the transaction costs—for both the DFI and the entrepreneur—of taking a prospective investment to the point where the decision to invest has been made, would also be prohibitive. Even if the costs were not an insurmountable barrier, there could be ethical questions around withdrawing DFI investment from the control group.

In principle, researchers might also hope to find a 'natural experiment' or exploit some sort of discontinuity in how DFIs select investments. But, to date, none has been found, to

[6] This paper focusses on evidence that might be obtained by researchers studying DFIs, not on what DFIs might be expected to report to shareholders and other stakeholders on a regular basis, but a similar point can be made: DFIs cannot measure their additionality, but they can be expected to show that they are investing in a way that is likely to be additional.

[7] Although the usefulness of RCTs is hotly debated: see e.g. Deaton and Cartwright (2017) and Imbens (2018).

[8] This figure is taken from the 2016 Annual Report of the Association of Bilateral European Development Finance Institutions.

[9] Evidence of additionality could plausibly be obtained by randomisation when a large numbers of small investments are being allocated: see McKenzie et al. (2017) for an example.

our knowledge. Hence, if quantitative evidence of additionality is to be found, it must be obtained from observational data that have been generated by economic actors going about their business.

2 DFIs and Additionality

This section discusses the concept of additionality, and describes how DFIs operate.

The simplest definition of additionality is: to make an investment happen that would not have happened otherwise (in the absence of the DFI's intervention).[10] DFIs sometimes refer to 'financial additionality' which denotes cases where the DFI offers finance to a project on terms that the market would not. We will avoid that term because whilst offering finance that private investors would not is necessary for additionality, it is not sufficient: a DFI can crowd-out private investors by offering project sponsors finance on more favourable terms than private investors would. We will use the terms 'investment' additionality or 'quantity' additionality to make it clear that we refer to an increase in the quantity of investment, relative to the counterfactual in which DFIs are not present.

Defined in this way, additionality is a partial-equilibrium microeconomic concept that relates to individual investments. General equilibrium outcomes, however, are of greater economic importance. If DFIs' investments are additional, we may expect the total quantity of investment to increase at the macroeconomic level when DFIs are active, but additional investments at the microeconomic level will not necessarily translate one-for-one into additional investment at the macroeconomic level. Mechanisms may exist whereby additional individual investments financed by a DFI cause some other investments to shrink or not take place, perhaps because there are supply-side constraints such as scarce human capital. Mechanisms may also exist so that the macroeconomic impact of an additional DFI investment is greater than one-for-one. If DFIs help create active capital markets and encourage entrepreneurs, they could indirectly cause other investments to take place (these are referred to as 'catalytic' effects). But despite some theoretical ambiguity about how additional individual investments will translate into changes in the overall level of invest-

[10] Independent Evaluation Group (2008) discusses the concept and lists the definitions used by various institutions (Box 2.1).

ment, the natural place to look for evidence of additionality is in the overall quantity of investment.

2.0.1 Other DFIs

When we define additionality as an investment that would not have happened otherwise we do not mean if another DFI would have otherwise done it. Suppose two DFIs exist, A and B, each of which invests \$50m annually in a series of small investments. For any individual investment, it could be true that if A had not done the deal, B would have (and vice versa) but it would be a mistake to say these DFIs have no additionality: if the total quantity of investment in the economy is \$100m higher thanks to the combined activities of the two DFIs, then they have both been fully additional. Which DFI took which deal is not important. The question that matters to, for example, a donor government considering injecting more capital into its DFI—and to researchers evaluating their impact—should not be substitution between DFIs but the margin between the public and the private sectors.[11]

2.0.2 Mobilisation

For the sake of simplicity, this paper describes crowding-out as a binary: each project is financed either by DFIs or by the private sector. In reality, DFIs sometimes co-invest with the private sector, and seek to 'crowd-in' private investors to projects they would not otherwise have supported. An investment can be additional and part-financed by the private sector. If the private sector would have financed the project without any DFI participation, then the DFI has partially crowded-out the private sector and has failed to be wholly additional. This nuance does not fundamentally alter any of the analysis in this paper and is henceforth ignored.

2.0.3 Development additionality

Increasing the quantity of investment is only one impact that DFIs intend to have on development. They also seek to achieve what some DFIs refer to as 'development' additionality

[11] An exception, where substitution between DFIs could be more worrying, would be if international DFIs crowd-out local DFIs (such as national development banks), which could be harmful for the development of local capital markets.

by changing the nature of investments so that they become more beneficial. DFIs will sometimes deliberately make an investment that the private sector would have made, with the intention of raising its quality from a development perspective. For example, an African cement manufacturer may consider taking money from a private equity fund that would then want it to focus on its strongest markets and maximise cash flow generation, with the intention of selling the business on again in a few years' time. Whereas a DFI may ask the business to open subsidiaries in poorer countries where the financial rewards will only emerge over the long run, but where the impact on development (creating jobs and reducing cement costs) would be higher. Development additionality could entail simply increasing a project's chances of success. For example, a project sponsor may have a willing private financier but believe that they have unrealistic expectations and would push the business in the wrong direction, to its detriment. Whereas a more patient DFI with more experience of the difficulties of creating sustainable businesses in developing countries could be more likely to see things through to a successful outcome. Some DFIs will admit that the majority of the projects that they invest in would have gone ahead in one form or another without them, and hence argue that their main contribution consists of their development additionality.

This paper focuses on investment additionality because it is the subject of most external commentary, is of the most political salience and is the cornerstone of the global financing for development strategy. Development additionality may be good for development, but it does not obviously contribute to the 'billions to trillions' strategy of leveraging additional private sector investment, which is the dominant narrative used to explain the importance of DFIs.[12]

Finding quantitative evidence of development additionality would be even more challenging than identifying investment additionality. It suffers from the same problem of an unobservable counterfactual, but the underlying concept is multi-dimensional and hence much harder to measure.

An important implication of this for those gathering evidence of additionality is that it would be helpful to differentiate between those investments where DFIs believe investment

[12] It is possible that DFIs could take an investment that the private sector would have made, and change things so it has a greater catalytic effect on subsequent private investments, but such arguments are rarely made.

additionality is present, and those which the private sector would have done but the DFI intended to deliver development additionality, before proceeding with the analysis.

2.0.4 Additionality and Market Failures

Additionality is not a sufficient test of positive development impact. Enterprises may produce outputs like jobs and taxes, typically thought of as good for development, yet not merit support from DFIs. In a well-functioning market, prices reflect the scarcity of inputs and the subjective valuations that consumers place upon final goods. Profitable enterprises create value, from a social point of view. If private investors are unwilling to support an enterprise because returns are too low, that could reveal that the enterprise would be destroying value from a social point of view. Using concessional finance to make such an investment happen could destroy value, despite the investment being 'additional' and despite it hiring workers and paying taxes. This is one reason why DFIs have traditionally tried to invest on commercial terms and to minimise the subsidy they confer on investees. Development impact only occurs when an investment is additional and social returns are greater than private returns (Warner, 2013). Social returns will deviate from private returns in the presence of market failures, including imperfect information, externalities and collective action (coordination) problems.

These points are important in theory, but may lack relevance in practice if market failures are so pervasive in developing economies that a modest degree of subsidy for investment can be justified if it creates any additional formal sector employment. Some economists argue that high levels of unemployment and underemployment, and large informal sectors, are symptomatic of widespread market failures and reason enough to provide some public support to marginal private investments.

2.1 How DFIs Work

The data generating process presented in the next section must correspond reasonably well to the reality of how DFIs invest for our results to have any validity. This section describes how the authors believe that DFIs operate.

DFIs do lots of things and this paper is concerned only with one of them: 'primary fundraising' intended to finance new economic activity, involving the installation of physical capital, investments in intangible capital, working capital to cover start-up losses, and so forth. DFIs also sometimes act as secondary markets, providing an exit for earlier-stage investors or refinancing existing loans, which implies a change of ownership or creditor without any new funds being raised for the enterprise itself. DFIs also attempt to help smaller businesses by providing earmarked lines of credit to local commercial banks and investing in private equity funds and other intermediaries that target small businesses, in an attempt to increase the supply of finance in markets where there is believed to be a shortage. For the sake of simplicity, the data generating process we develop in this paper describes direct investments and does not attempt to model such activities, although they are essentially more convoluted ways of increasing the quantity of investment.

DFIs have a 'demand-led' business model, which means that they rely on others—known as 'project sponsors'—to come up with investment ideas and come looking for money.[13] From the point of view of the researcher, investment opportunities are unobservable — we can only observe opportunities that succeed in receiving an investment. The magnitude of an investment (the sum invested) is largely determined by the nature of the underlying enterprise: if the project sponsor sees an opportunity in manufacturing shoes then they must raise whatever sum of money is required to cover the costs of getting a shoe factory up and running (although of course the size and type of shoe factory is subject to negotiation with financiers).

The typical DFI investment is agreed with the project sponsor after confidential bilateral negotiations, with DFIs sometimes acting as part of a consortium, which may also include private financiers. The project sponsor may be in talks with other DFIs and private financiers, but is under no obligation to divulge the contents of those discussions to other parties. The project sponsor may care about many things, but somewhere towards the top of the list is obtaining finance on the most favourable terms. The risk of crowding-out arises because DFIs are typically more attractive to project sponsors than private financiers. This does not necessarily mean that DFIs' pricing is always more favorable; they also offer a range of non-financial benefits, such as political protection, which can make them more attractive to

[13] DFIs are able to be proactive to a certain extent, and can allocate grants to cover early-stage project development costs.

sponsors.

As a result, whether an investment is additional is not known with certainty by the DFI, who does not know if there are willing private financiers waiting in the wings. The project sponsor may not know either, as they may enter negotiations with DFIs before they have exhausted all private possibilities. Sometimes DFIs may know that private financiers are bidding against them, but that does not mean that if the DFI withdraws, the private financiers will necessarily bring the project to fruition.

The traditional DFI model is based on investing on commercial terms; concessional finance is also available, but most DFIs draw a sharp distinction between concessional financing and their main business. Some DFIs set their prices with reference to estimated market prices, others have a cost-recovery model. Because market prices often imply returns above cost-recovery, cost-plus prices are sometimes regarded as concessional by the market-down pricers. This is a point of disagreement in the industry. Investing on commercial terms does not mean exactly mimicking private investors; DFIs would have little reason to exist unless they do something the private sector does not. But within the constraints of their business model and the project's financing needs, they still attempt to drive a hard bargain with project sponsors.[14]

DFIs usually have a mandate from the shareholders to be self-financing, some even pay dividends, but many DFIs also believe that investing on commercial terms is good for development. DFIs mostly do not want to create businesses that are reliant on subsidised finance to survive. Their goal is to create sustainable businesses that create social value. That will not happen if the businesses that DFIs invest in collapse the moment they are forced to refinance at market rates. Concessional finance also risks distorting markets by, for example, allowing subsidised firms to drive more productive firms out of business. In some sectors, such as off-grid solar, subsidies are used routinely and so the prices charged to consumers do not reflect unsubsidised cost recovery for the suppliers. Hence it is not clear whether there is a sustainable business model. In cases such as these, subsidies may help create new markets if, for example, they accelerate 'learning by doing', technological progress and economies of scale, until viable businesses emerge.

[14] A distinction can be made between returns at the project level and net returns after central overheads. A DFI may demand the same returns at the project level as commercial investors but tolerate lower net returns after central costs, or more variable returns.

From the project sponsor's perspective, not everything about DFIs is more appealing than private finance—DFIs are more troublesome to do business with because they impose higher environmental, social, and governance standards; they ask investees to report development outcomes, which is costly; and they may also interfere with corporate strategy. By pricing on close-to-market terms and not offering supplementary benefits, DFIs may thereby minimise crowding-out because project sponsors will choose private finance if they can get it.[15]

Because DFIs want to behave like commercial investors, their investment teams are remunerated in similar ways to commercial investors: on the basis of deal volume and returns. The risk of crowding-out arises because DFIs' investment teams therefore have an incentive to go after the same investment opportunities as private investors. Since additionality is not observable, the indicators of development impact that DFIs care about, jobs created, taxes paid and so forth, can be reported even though an investment is not additional. DFIs as institutions can also face incentives to 'get money out the door' and claim success accelerating investment in developing countries. Typically, a DFI's investment teams will submit their proposals to an investment committee, where a view will be formed on the likely additionality of the deal (and its suitability in other respects). A successful DFI must manage these conflicting incentives.

Even with the best intentions, we should expect DFIs to make errors. The risk of making the wrong decision about additionality is not one-sided: DFIs may do deals that are not additional, but they may also turn away deals that would have been additional. From the perspective of maximising development impact we may want DFIs to err on the side of assuming additionality, on the basis that occasionally crowding-out private financiers has a smaller social cost than failing to support worthwhile additional investments.

Another reason why DFIs may fail to be wholly additional could be because they cross-subsidise investments across their portfolios, and are thereby able to fund a larger absolute number of additional projects and have a greater total impact on development. To see how this might work suppose for the sake of simplicity that the threshold (minimum) expected return demanded by private investors is 10%, and 5% by DFIs. There are a finite number

[15] Hainz and Hakenes (2012) shows how development banks can set their prices to screen out project sponsors that should not receive public support.

of investment opportunities with expected returns in the 5-10% range. Let's say there are 100 such projects, giving an average return of 7.5%. Instead of only choosing those 100 investments and being 100% additional, DFIs can achieve the same average returns by using investments with expected returns above 10% to cross-subsidise some with expected returns below 5%, with the result that they invest in more than 100 additional projects at the cost of also displacing private investors from some projects with returns above 10%. If DFIs are using such an approach the proportion of additional investment in a DFI's portfolio would depend on the distribution of expected returns across investment opportunities and the nature of its business model and financial targets.

3 The Data Generating Process

This section describes the data generating process (DGP) that we use to create simulated datasets on which we test whether standard empirical techniques can identify additionality of DFI investment in observational data. We use the same DGP for analysing both cross-country and firm-level identification strategies, although not all features of the DGP will be relevant in each case.

Although this DGP will necessarily be highly stylized and simplified, it is intended to be realistic, in the sense of capturing those elements that are necessary to our analysis. It reflects our understanding of how DFIs operate, as articulated in the previous section. Besides providing a testing bed for empirical methods, we also think it is useful to translate our verbal description of how DFIs operate into a quantitative process. The debate over additionality may be sharpened if disagreements are also expressed in terms of different statistical processes, which may make it easier to see how consistent alternative theories are with the data.

Rodrik (2015) argues that unrealistic assumptions are inevitable in any model, but only present a problem when they are critical to the results of the analysis. Our DGP ignores many features of reality that might be important in other contexts, but which we believe are not critical to our results. For example, the number of investment opportunities in a country has something to do with its size (population) but modelling countries of different size would have no bearing on the qualitative conclusions we reach. We have tried to make the DGP as simple as possible, but no simpler, as the saying goes.

We take the approach of asking whether it is possible to identify additionality under relatively benign conditions. We could make our data generating process more realistic in a number of ways which would make identifying additionality even more difficult. We hope not to have excluded anything that matters in reality and would make identifying additionality easier.

The DGP does not include any mechanisms wherein an additional investment, in the microeconomic sense, could crowd-out investment opportunities at the macroeconomic level. Each additional individual investment will translate one-for-one to additional investment at the macroeconomic (country) level. The converse is not necessarily true: DFIs may crowd-out private investors from a given investment opportunity but may increase the total quantity of investment if displacing private funds from one investment frees money to be invested elsewhere.

In each period every country generates a fixed number of investment opportunities with an associated expected return, which is not observable by researchers. We assume investments are all of the same size (normalised to 1). The expected return of each investment opportunity is a stochastic function of an associated variable we call 'project characteristics' which we conceive of as summarizing all observable data about the project, such as the sector and geography of the project, and management's track record. Although we refer simply to expected returns, we have in mind risk-adjusted returns. The stochastic nature of the expected returns function means that some unpromising projects, on the basis of their observable characteristics, possess high expected returns, whilst some seemingly attractive projects have low expected returns. We shall test whether it is possible to use observable project characteristics to produce a good estimate of additionality.

In reality, expected returns may depend on the match between the project sponsor and the investor, and the nature of the project will be subject to negotiation with each investor. Rather than being a known quantity, project sponsors and potential investors will each form their own subjective expected returns, which will differ from the project's true expected return. Introducing these elements to the DGP may make it more realistic, but we believe it would not alter the conclusions, so we treat expected returns as a single number privately observed by all investors.

We suppose that there are three varieties of countries: emerging markets (e.g. Vietnam), frontier markets (e.g. Tanzania) and those beyond the frontier (e.g. Chad). The variable we refer to as 'project characteristics' that is the basis of each investment's expected returns is drawn from a different distribution for each type of country, with mean returns being high, medium, and low, respectively. Each country draws the same number of investment opportunities, each period. Because investors have minimum expected return thresholds, the distribution of expected returns determines the number of viable projects in each country. The quantity of investment varies across countries because the number of opportunities that are bankable (have sufficiently high returns) will vary across countries according to type. Our DGP will create, on average, many viable investment opportunities in emerging markets, an intermediate quantity in frontier markets, and few bankable projects in countries that are beyond the investment frontier.

The investment frontier moves across countries over time, so in our DGP in each period there is a probability that a country will change its type. These probabilities are defined by a fixed transition probability matrix. We assume the world is developing, so the probabilities of moving up the hierarchy are higher than the probabilities of regress. Variation over time is potentially important because some empirical methods can account for unobserved factors (such as the returns probability distribution) if they do not vary over time.

Our choice of probability transition matrix to govern the evolution of country types over time implies that the quantity of investment opportunities increases over time (on average). We model the DFI budget as a random walk with drift; if the drift we choose is positive then the DFI budget will trend upwards, introducing a correlation with the total quantity of investment even when DFIs are having no impact.

It may be worth emphasising that the country-type transitions and the DFI's budget are the only dynamic elements of our DGP (meaning the past affects the present). Everything else happens within each period in isolation. The stochastic processes that determine investment opportunities have no serial correlation. We do not know if that is realistic or not, but it may matter for the performance of some econometric methods where serial correlations in errors can cause problems.

DFIs and private investors each have a minimum (threshold) expected return, below which they will not invest. This is a simplification but captures the basic idea that there are some projects that the private sector will invest in, and others that they will not, and that DFIs may invest in either type. If DFIs have lower threshold returns than private investors, they can be wholly additional if they finance only projects below the private threshold. But DFIs may crowd-out private investors by financing projects above the private returns threshold.

Both DFIs and the private sector have a 'budget' each period. An exogenous DFI budget will allow us to examine the performance of 'instrumental variable' strategies, that rely on a source of exogenous variation to separate correlation from causation.

In reality, DFIs and private investors may have some constraints on how much they can invest each period, but they do not have fixed and exogenous budgets. The spending power of a DFI—its 'budget'—has to do with new capital contributions from shareholders, the sale of previous investments, and in some cases decisions to borrow on capital markets. The idea of a fixed budget for private sector investors would be unusual: a standard assumption in a 'small open economy' model would be that there is a global capital market, with a minimum required risk-adjusted return, and that all projects with a risk-return profile above that threshold will be financed.

This may matter, because if private budgets are constrained so that investment opportunities exceed the supply of finance, then DFIs may crowd-out private investors from individual projects yet still increase the overall quantity of investment, by freeing-up private funds to be invested elsewhere. Perhaps in practice private investors do allocate fixed sums to funds that invest in emerging and frontier markets, so the idea of a constrained supply of private finance may be more realistic in less developed countries, at least in the short-run. Our DGP grants us the ability to implement this possibility, but all the results we have chosen to report in this paper are based on the 'small open economy' assumption that private investors are able to finance all bankable investment opportunities.

In our main specification, when the quantity of eligible investment opportunities exceeds the available budget, we assume projects are selected at random. This does not imply that investors are acting at random but rather that, conditional on projects offering acceptable expected returns, whatever the process is that determines which projects investors choose, it looks random from the outside.

Private investors will finance projects with expected returns above a minimum threshold. DFIs will select projects from a set of eligible projects defined by a lower and an upper bound of expected returns, until their budget is exhausted. We introduce this upper bound to allow us to vary the probability that a DFI will crowd out private investors. If the DFI upper bound is beneath the private threshold, they will be wholly additional (there is no overlap between the two sets). The higher the DFI upper bound is, the more DFIs encroach upon private sector territory and the probability of DFIs taking a private project will rise. We assume that DFIs crowd-out the private sector whenever they are competing for the same investment.

The assumption that DFIs select projects at random from investment opportunities in their eligible set is crucial, because it means that DFIs' investments are positively correlated with the quantity of investment in each country even if DFIs entirely crowd-out private investors. Countries with more investment opportunities account for a larger share of the eligible set and hence are more likely to be chosen.

To show how this assumption affects patterns in the data, we introduce an alternative DFI project selection process. To do that, we will exploit the 'project characteristics' variable. This variable measures expected returns with error. Some projects with 'low' project characteristics will have high expected returns and vice versa. As an alternative project selection process, DFIs will rank investments within their eligible set from low to high according to project characteristics, and invest in ascending order until the budget is exhausted. If the DFI budget is sufficiently small, so that selection within the eligible set is important because the entire set does not receive investment, DFI investments will now be negatively correlated with the total quantity of investment in each country, because low project characteristics are more common in countries beyond the frontier, where the overall quantity of investment is lower. This alternative selection process could be interpreted as DFIs having a mandate to do deals in difficult places, so they prioritise those that superficially look like deals the private sector would avoid.[16]

[16] A negative association between DFI investments and total investments across countries can also be obtained when projects are selected at random by setting low and narrow bounds on the eligible set, because in our data generating process lower-return investment are relatively more common in countries with lower

3.1 Multiple DFIs

To fully investigate the performance of a supply-push instrument in this setting, we extend our data generating process to multiple DFIs. Different DFIs have different country preferences, and this contributes to the persistence of DFIs' investments over time. To avoid having to create an ad hoc method for allocating investments to competing DFIs we exploit the fact that DFIs often collaborate by co-investing in the same project. We introduce 'country preferences' for each DFI, which are then weighted and used to determine what share of each investment in a given country is taken by which DFI. This method allows us to leave the basic mechanism for selecting investments at random from the eligible set unchanged. The weight placed on country preferences can be adjusted between zero and an arbitrarily large number (infinity). When the weight is zero, all investments are divided up equally between DFIs and there is no meaningful change from having a single DFI; when the weight is large (infinity) the DFI with the strongest preference takes the entire investment. We generate country preferences for each DFI-country pair as a random number between 0 and 1 drawn from a uniform distribution, and then transform these so that they sum to 1 across DFIs. Hence when the weight on country preferences is set to 1 the share that each DFI takes of an investment in a given country is equal to its numerical preference for that country.

We assume the following variables are available to the researcher: individual DFI budgets, the number of investments that DFIs and the private sector make in each country in each period, and the variable that summarises observable project characteristics. In the cross-country regressions these variables are either a sum (investments) or the mean (project characteristics) at the country level. In the firm-level regressions, we use the individual investments. The expected return of each project and the private sector's threshold are not observable (otherwise it would be easy to see which investments are additional).

overall levels of investment. If the size of the DFI budget is small relative to the size of the eligible set of investment opportunities, then the ranking of opportunities within that set matters more; if the budget is large then the boundaries of that set matter more.

3.2 Formal Description

This section sets out our data generating process more formally and specifies the default values of parameters that we use. We later vary the values of some parameters to explore the performance of econometric techniques under different circumstances.

There are three types of country, indexed by c. Project characteristics are drawn from a normal distribution specific to each country type:

$$pc \sim \mathcal{N}(\mu_c, \sigma_c^2) \tag{1}$$

Standard deviations σ_c are set to 1 by default. μ_c is set to $[0, 2, 4]$ for the three country types by default, respectively. Expected returns are the sum of project characteristics and a mean zero normally distributed random variable:

$$er = pc + e \tag{2}$$

Where

$$e \sim \mathcal{N}(0, \sigma_e^2) \tag{3}$$

The default value for σ_e is 1.

We construct datasets with $nC = 80$ countries, which is a generous estimate of the number of countries for which a researcher may obtain data on DFI investment and other necessary variables. We initiate the DGP with half the countries as type 1 (low average returns), a third type 2 (medium average returns) and a sixth type 3 (high average returns). The probability transition matrix that governs how country types evolve over time is:

$$P = \begin{bmatrix} .85 & .10 & .05 \\ .05 & .85 & .10 \\ .05 & .05 & .9 \end{bmatrix} \tag{4}$$

The rows correspond to type at the start of the period, the columns the type at the end. So for example a type 1 country has a 85% chance of staying type 1, a 10% chance of becoming type 2, and a 5% chance of becoming type 3, and so on for the other types.

In the default set-up, we have $T = 20$ periods, and $nI = 50$ investment opportunities are available in each country-year. Hence the total quantity of investment opportunities in each period is $nC * nI$.

The private sector deems a project worthy of investment if its expected return exceeds the lower bound ps_{min} (default: 2). The parameters we have chosen imply that on average just over 5% of investment opportunities in type 1 countries are bankable, 50% in type 2 countries and just over 90% in type 3.[17] The bounds of the DFI eligible set will be varied in different experiments, to generate situations with zero and full additionality. The lower and upper bounds that delineate the set of eligible DFI investments are denoted dfi_{lo} and dfi_{hi}, respectively. To get zero additionality with our default choice we set $dfi_{lo} = 2 = ps_{min}$. To create a situation with full additionality, we set $dfi_{lo} = 0$ and $dfi_{hi} = 2$, so that there is no overlap in the sets of projects the DFI sector and the private sector are potentially interested in. Each investment that is chosen by either DFIs or private investors is coded with a 1 and the total quantity of investment in each country-period is simply the sum of these.

For example, if the DFI eligible set is [0,2] so that DFIs are wholly additional, just over 40% of investment opportunities in countries type 1 and 2 are eligible, but only around 8% in type 3. Within these sets, the medians of the project characteristic variable are approximately $[0.4\ 1.6\ 2.7]$ - recall that $er = pc + e$ so that an opportunity with $pc = 2.7$ may have $er < 2$. Hence under the alternative DFI project solution procedure which ranks by project characteristics and chooses the lowest first, when the DFI budget is substantially smaller than the eligible set, investments will be concentrated in type 1 countries.

In the first period, by default we set the DFI budget DB_1 equal to 10% of the total number of investment opportunities in that period (i.e. without regard for expected returns): $DB_1 = 0.1 * nC * nI$ (equal to $0.1 * 80 * 50 = 400$ for default parameter values). This budget then evolves according to a random walk with drift:

$$DB_t = DB_{t-1} + drift + \eta \text{ with } \eta \sim \mathcal{N}(0, \sigma_{db}^2) \tag{5}$$

Default parameter choices are $drift = 0$ and $\sigma_{db} = 0.2 * DB_1$. We avoid the possibility of negative DFI budgets by setting a lower bound of 20 for DB. That is, the evolution of DB

[17] These numbers and others that follow were taken from a very large draw of the simulated data.

is given by equation (5) but if $DB_t < 20$ then we replace it by $DB_t = 20$ and continue from there. By default we assume that the private sector budget in each period exceeds the total number of investment opportunities.

In our default set-up, we consider 3 DFIs ($nD = 3$). DFI-specific preferences for countries are time-invariant and generated as follows:

$$p_{d,i} = \frac{\pi_{d,i}}{\sum_{d=1}^{nD} \pi_{d,i}} \text{ with } \pi \sim U(0,1) \tag{6}$$

where d indexes DFIs and i indexes countries. $p_{d,i}$ is between 0 and 1; the DFI with the highest preference takes the largest share of each investment, to an extent determined by the weight ϕ. The share of each investment in country i taken by DFI d is then given by:

$$S_{d,i} = \frac{{p_{d,i}}^{\phi}}{\sum_{d=1}^{nD} {p_{d,i}}^{\phi}} \tag{7}$$

4 Evidence from Country-Level Data

The scarcity of country-level data for DFI investment has meant that cross-country empirical analysis is limited. Dirk Willem te Velde and co-authors have produced a series of quantitative research papers, based on investment data the authors obtained from DFIs. Massa et al. (2011) report that a 1 percentage point rise in the investments of two large DFIs (the IFC and EIB) as a share of recipient GDP is associated with a 0.8 percentage point rise in gross capital formation, as a share of GDP. More recently, however, Massa et al. (2016) report positive and statistically significant estimated coefficients on the investments of a subset of individual DFIs in gross capital formation regressions, but no statistically significant estimated coefficient when the investments of DFIs are combined. For some individual DFIs the estimated coefficients are large (approximately 6) but disappear once time dummies are added to the regression; for others the coefficients are in the range $1-3$ and retain statistical significance across specifications. Some specifications control for other development finance flows (aid and FDI) but they include no variables that might serve to predict the level of investment (gross capital formation) across countries and over time—in the context of our data generating process, that translates as not controlling for observable project characteristics. The authors used fixed effects and random effects regressions, with no instrumentation strategy to adjust for the non-random allocation of DFI investments.[18]

4.1 Lessons from Simulated Data

The most fundamental, and unsurprising, insight from our data generating process is that the nature of the investment selection mechanism used by DFIs can create either positive or negative correlations with total investment across and within countries, whether DFIs are additional or not. If DFIs have zero additionality, and pick investment opportunities at random from the same pool as private investors, there is a strong positive correlation between DFI and total investment. In countries with more plentiful investment opportunities, offering returns over a minimum threshold shared by both DFIs and private investors, there will both be more investment by DFIs and more overall investment.

Conversely, DFIs can be fully additional and yet DFI investment may exhibit a negative correlation with overall investment, if their investment selection mechanism causes their investments to be concentrated in countries with lower levels of investment. This can happen in our data generating process either because the choices of eligible set and distributions of expected returns in country types results in DFI investments being concentrated in countries with less overall investment, or because DFIs are ranking opportunities by project characteristics and selecting the lowest. This could be loosely interpreted as having a mandate to target investments in the most challenging, least appealing markets. For example, Massa et al. (2011) report that the British DFI, CDC Group, is unusual in skewing its portfolio more towards poorer countries; the gross capital formation regressions in Massa et al. (2016) estimate a large negative correlation between CDC investments and total investment across countries.

These correlations are strongest across countries, but they also exist within countries in time-series variation. Panel regressions with fixed effects can absorb the variation in type across countries but even with country types fixed over time, there is random variation in the set of investment opportunities which will generate spurious within-country correlations between total and DFI investment. In our data generating process, country types

[18] Some of the regressions in these papers use GMM but not the capital formation regressions.

evolve according to a probability transition matrix intended to mimic the investment frontier moving across countries over time, so as countries transform into emerging markets that offer greater investment opportunities, DFI investments will be correlated with periods of lower or higher overall investment, as reflects their investment selection method.

Good control variables can reduce bias, but cannot prevent a misleading result. In our simulated data, the variable 'project characteristics', which is the within-country average of the project characteristics of individual investment opportunities, is an excellent predictor of the overall level of investment. When it is included in regressions, the size of the bias is reduced, but some bias remains. There is a gap between observable project characteristics and privately observed expected returns. Since investors base their decisions upon the latter, the inclusion of project characteristics in regressions does not fully prevent misleading results (false positives or negatives).

In applied research, it is plausible that econometricians might include a set of variables that collectively proxy for the underlying determinants of investment demand within countries, but they are never going to obtain variables that predict investment demand as accurately as the project characteristics variable in our simulated data. So, in some iterations of our simulations, we add some well-behaved measurement error to this variable. After taking the country mean of project characteristics $\bar{pc}$ we add a zero-mean normally distributed error term:

$$\widetilde{pc} = \bar{pc} + \nu \tag{8}$$

Where

$$\nu \sim \mathcal{N}(0, \sigma_m^2) \tag{9}$$

The larger is σ_m^2, the less bias is removed by the inclusion of project characteristics in our regressions.

The basic model we estimate is:

$$I_{it} = \beta dfi_{it} + \gamma \widetilde{pc}_{it} + z_t + w_i + u_{it} \tag{10}$$

where I_{it} is the total investment received by country i in period t, dfi_{it} is the amount of DFI investment and $\widetilde{pc}_{it}$ the average project characteristics in that country-period after adding measurement error ν as specified in equation (8). Not all elements are present in each regression; sometimes $\widetilde{pc}$ is excluded for instance. The fixed effects estimator absorbs country-specific intercepts w_i, while pooled OLS leaves them as part of the composite error term $w_i + u_{it}$. As is standard in panel data regressions, we include time dummies z_t to account for common shocks over time that affect all countries equally.

Because our data generating process is stochastic and our sample is relatively small, the estimated coefficients in our regressions are subject to natural random sampling variation. Hence, we report the average estimated coefficients from 1000 iterations of our data generating process and also the standard deviation of those coefficients across iterations.

Table 1: OLS and fixed effects regressions

True β	0	0	0	0	1	1	1	1
Selection	Rand	Rand	Rand	Rand	Rank	Rank	Rank	Rank
PC	Excl	None	Some	High	Excl	None	Some	High
OLS β	3.13	0.24	1.02	1.94	-1.82	0.95	0.29	-0.57
Std Dev	0.86	0.06	0.17	0.28	0.62	0.05	0.09	0.14
FE β	2.03	0.23	0.91	1.51	-0.85	0.94	0.29	-0.31
Std Dev	0.33	0.06	0.16	0.22	0.22	0.05	0.13	0.16

In all regressions in the first four columns, the null of zero additionality is rejected at the 5% level in 100% of iterations. 'Selection' indicates the method that DFIs use to choose investments from their eligible set: at random, or by ranking by project characteristics and picking the lowest first. 'PC' indicates whether the variable project characteristics is included and if so how much measurement error (noise) has been added to it. In the first four columns DFIs have zero additionality, in the second four DFIs are fully additional.

Table 1 reports the mean estimated coefficients for ordinary least squares (OLS) and fixed effects (FE) regressions, where the latter estimates derive from variation within countries. Fixed effects estimates control for unobserved country characteristics when they do not change over time, but our country types, and hence the distribution of investment opportunities, evolve over time.

These results show how it is possible to obtain positive estimates when DFIs have no additionality (the first four columns) and negative or close to zero estimates when they are fully additional (the second four columns). Including the variable 'project characteristics' brings the estimated coefficients closer to their true values, but even when there is no measurement error (noise) added to that variable, there is still substantial bias. We add measurement error to the project characteristics variable because in reality any regression that tries to explain the level of investment across countries will have a much smaller R-squared, since no researcher will have access to variables that collectively predict the quantity of bankable investment projects as accurately as our project characteristics variable. In column 2 with no measurement error, the mean R2 in the OLS regressions is 0.98. The standard deviation of the error term added to *pc* is 0.5 in the columns headed 'some' and 1 in those headed 'high'. Even with noise added to *pc* these regressions still explain a higher proportion of the variation in the data than researchers may typically expect: in column 4 with noise labelled 'high' the mean R2 across the OLS regressions is 0.84.[19]

The 'false-positive' result, when DFIs have no additionality and select projects at random from their set of eligible investments, is caused by their demand-led business model: when a country gets a good draw of investment opportunities it will attract more investment from both DFIs and the private sector. We have illustrated a 'false-negative' result under full additionality when DFIs select the least-appealing projects first, as judged by their observable characteristics. Equally large negative correlations can be generated by choosing an eligible set that contains investment opportunities with expected returns that are more frequent in countries where the overall quantity of investment is lower.[20]

Our results show how the inability to control accurately for the determinants of investment increases the omitted variable bias which loads onto the DFI coefficient. The bottom line is that without knowing how DFIs select their investments, estimated coefficients in cross-country investment regressions tell us nothing about additionality.

5 Internal Instruments

The previous section illustrated how the estimated coefficient on DFI investment in a cross-country investment regression can be of either sign, depending on how DFIs select their investments, and regardless of whether DFIs are fully additional or have zero additionality. That happens because DFIs' investments are 'endogenous', because they are correlated

[19] Although the R2 is already high in our regressions in cross-section, at around 0.7, because DFI investments are highly correlated with overall investment. Adding PC has a larger impact on the within-R2 in fixed effects regressions.

[20] This latter possibility could be regarded as an artefact of our DGP, which uses a symmetric distribution to generate investment opportunities. As a result, when investments with higher expected returns are plentiful, as in type 3 countries, investments with low expected returns are rare. One might think that in reality, bad investments are plentiful everywhere, if one cares to look for them. Or perhaps project sponsors everywhere only seek funding when they have an opportunity they expect to be profitable, then in the eyes of foreign investors macro risks shift all expected returns down in frontier markets. Because we do not know the true distribution of expected returns across countries, we regard a preference for superficially unappealing projects as a more realistic potential mechanism for generating false negative results.

with a variable that affects the dependent variable but is not explicitly controlled for in the estimated regression (in our case, the number of eligible investment opportunities). To identify additionality—estimate the causal impact of DFI investments—exogenous variation is required, meaning variation in DFI investment that is uncorrelated with the error term.

A popular approach to dealing with endogeneity in cross-country regressions is to use an estimator that exploits 'internal instruments' to isolate exogenous variation in DFI investment. The most popular of these is the system GMM estimator developed by Arellano and Bover (1995) and Blundell and Bond (1998), which we implement using the xtabond2 software provided by Roodman (2009a, 2015). The 'internal' instruments are lagged (historical) values of the variables suspected of endogeneity. The basic idea is that historical values of variables can be used to predict future values, but will not be correlated with the contemporaneous factors that are the source of endogeneity (such as the error introduced by how the availability of investment opportunities differs from observable project characteristics).[21]

If the instruments generated by system GMM are valid (uncorrelated with the error terms of the relevant equations), the estimator is consistent. Roughly speaking, this means that it correctly identifies the causal effect of interest in large enough samples. The validity of the internal instruments is not a given, though. In a standard GMM configuration, for instance, instrument validity requires the absence of serial correlation in u_{it}. It also depends on an 'initial conditions' restriction that is, again, not guaranteed to hold in practice (see Roodman, 2009b, for an excellent discussion). Moreover, the internal instruments might only be weakly correlated with the variables they are meant to instrument, and this can lead to large biases (Bun and Windmeijer, 2010). Unlike in the case of standard instrumental variables estimation (see later), no straightforward weak instrument test is available to researchers, so the latter problem is often left undiagnosed.[22]

[21] More precisely, the difference GMM estimator (Holtz-Eakin et al., 1988; Arellano and Bond, 1991) takes the first difference of equation (10) —subtracts one year from the next—to remove w_i, then uses suitably lagged levels of the right hand side variables as instruments. The system GMM estimator we focus on here further adds equations in levels, instrumented by lagged differences of variables. Bond (2002) and Roodman (2009a) provide excellent guides to these estimators.

[22] Kraay (2015) shows how weak instruments can lead researchers astray in cross-country regressions and suggests several ways of testing for weak instruments and of adapting inference accordingly, but Windmeijer (2018) shows some of these tests are invalid.

Our data generating process is a benign setting for the system GMM estimator, because the stochastic process that generates investment opportunities is drawn afresh each period. As a result, this element of the data generating process does not introduce serial correlation into the error term that might otherwise confound a GMM estimator. Nonetheless, in the models we estimate with system GMM we include a lagged dependent variable, in order to mimic the 'standard' specification to which system GMM is most often applied.[23] We treat the lagged dependent variable and the project characteristics variable as predetermined, and the DFI investment variable as endogenous.[24] To avoid overfitting (Roodman, 2009b), for each variable we restrict the number of lagged levels used to instrument the differenced equation to 4, and we also collapse the instrument matrix. We report estimates from the two-step GMM estimator.

As table 2 shows, when the data generating processes includes realistic amounts of drift in the global DFI budget and noise in the project characteristic variable, the GMM estimator consistently produces misleading results. These results are from the DGP with zero additionality. When there is no measurement error or drift, the estimator performs well on average (the mean β is close to zero), although the variance of estimated coefficients is quite high across draws and the null of zero additionality is rejected at a 5% significance level on 10% of occasions. But when measurement error is introduced to the project characteristics variable, there is upward bias on the estimated β on DFI investments. When the quantity of measurement error is 'high'—a normally distributed error term with standard deviation of 1 added to a variable with mean values $[0\,2\,4]$ in the three country types respectively —then the mean estimated β is close to 1 and zero additionality is rejected at the 5% level in 72% of iterations. The introduction of drift reduces the bias and reduces the variance of estimated β across iterations, thereby somewhat increasing the frequency with which the zero additionality null is rejected. It is quite likely researchers would conclude that DFIs are fully additional, on the basis of such regressions.

GMM estimators are complicated and have many moving parts. The reported bias could be due to invalid instruments or due to weak instruments, or a combination of both, as any

[23] In practice, this specification is the most commonly employed with system GMM because the inclusion of a lagged dependent variable typically strongly reduces the degree of serial correlation in the error term, thereby making it more likely that at least one necessary assumption for consistency is satisfied.

[24] These choices affect exactly which lags are used as instruments. See Roodman (2009a) for details on these and other specification choices.

Table 2: GMM produces false positives in the presence of measurement error

Noise Drift	None No	Some No	High No	Some Yes	High Yes
Mean β	0.05	0.38	1.13	0.37	1.07
Std Dev	0.14	0.47	0.81	0.41	0.66
% reject H0	11	35	72	42	80
m1	100	100	100	100	100
m2	90.9	88.7	87.6	87.9	86.3
Hansen	80.6	33.6	20.5	27.2	15.8
Sargan	79.3	25.8	11	21.5	9.5

This table reports the average coefficients on DFI investments from system GMM regressions, together with their standard deviation, across 1000 replications, and the percentage that reject the null of zero additionality at the 5% level. The first row indicates whether there is measurement error added to the variable 'project characteristics' and the second row whether there is a drift term in the random walk that determines DFIs' budgets. The bottom four rows give the percentage of iterations that 'pass' specification tests that researchers commonly run.

bias introduced in the GMM estimator by invalid instruments could be exacerbated by instruments being weak. The lack of readily available tests for instrument strength in this context further complicates matters, as does the use of many different instruments, which makes it more difficult to diagnose exactly which instruments are invalid. Hence, without further detailed analysis, which is beyond the scope of the current paper, it is difficult to pin down exactly what is the main source of bias. To narrow down the possibilities, the table reports the percentage of replications in which a small number of diagnostic tests is 'passed' (i.e. do not suggest problems in applying the estimator) at a 10% significance level. The $m1$ and $m2$ tests do not point to problems in the vast majority of replications, indicating that serial correlation in u_{it} is not the source of the bias.[25] When we find bias, the Sargan and Hansen tests of instrument validity only 'pass' (i.e. do not reject instrument validity at a 10% significance level) in a small fraction of replications. This is a strong indication that at least some of the internal instruments are not valid.[26] What is clear, however, is that, once our DGP contains elements which are very easy to believe would also be present in real data (including, for instance, the availability of imperfect control variables to reduce omitted variable bias) then the GMM estimator produces very misleading results.

[25] Recall that instrument validity in system GMM requires a lack of serial correlation in u_{it}. If this is so, Arellano and Bond's (1991) $m2$ test should not reject the null hypothesis of no second-order serial correlation in Δu_{it}, while the $m1$ test should reject the null of no first-order serial correlation in Δu_{it}. Both of these are the case in a large fraction of replications.

[26] Difference-in-Hansen tests that probe the validity of subsets of instruments (results not reported) appear to point to problems across the board, especially with the instruments for the levels equations and the instruments based on lags of DFI investment. But we reserve a more detailed analysis for a future date.

6 External Instruments

The alternative to estimators that exploit internal instruments, such as GMM, is to look for an external instrument that will allow the researcher to separate correlation from causation. A valid external instrument is something (known as an instrumental variable) that is correlated with the variable of interest, in this case investments by DFIs, but is uncorrelated with the error term and has no independent effect on the outcome of interest, in this case the overall quantity of investment in an economy. Valid instruments are notoriously difficult to find in the context of development (Bazzi and Clemens, 2013) but a natural candidate would be a 'supply-push' instrument (also known as a shift-share or Bartik instrument, after Bartik (1991)). Variants of this instrument have been used by several recent papers in the aid effectiveness literature, see for example Werker et al. (2009); Nunn and Qian (2014); Dreher and Langlotz (2015). We construct our instrument as in Temple and Van de Sijpe (2017).

The basic idea is that the budgets of development agencies fluctuate for reasons that are unconnected to changes in the investment climate in recipient countries and that development agencies have persistent preferences for some recipient countries over others. That means that when a DFI's budget increases, some recipient countries will experience a larger increase in DFI investment than others for reasons that should be uncorrelated with a recipient's current circumstances.

As we discuss below, however, application of the supply-push IV to the current context is likely to encounter some difficulties that could result in bias, caused by the 'demand-led' nature of DFIs' business models. It is reasonable to suppose that DFIs have persistent preferences for some countries over others, caused by historical ties or explicit strategy decisions, and that the quantity of total investments DFIs make will respond to whatever investment opportunities arise in those countries. If the countries they favour experience an improvement in the investment climate, DFIs are likely to respond by increasing their total investments. DFIs do not really have a predetermined budget to spend each year, as aid agencies that largely disburse grants do. Rather they have access to financial resources which, to an extent, they can 'draw down' in response to demand. Hence a DFI's 'budget' (the total quantity it invests in any given period) is not really exogenous to changes in circumstances in countries a DFI has a preference for, as represented by its pattern of investment over some initial period. As we show below, this mechanism contaminates the supply-push instrument, so that it may yield a biased estimator of DFI additionality.

6.1 The Instrument

The supply-push instrument looks as follows:

$$dfiIV_{it} = \sum_{d=1}^{nD} s_{i0}^{d} D_{dt} \tag{11}$$

where s_{i0}^{d} is the share of DFI d's total investment budget that country i receives over an initial period that is excluded from estimation, and D_{dt} is the total investment made by DFI d in period t. The number of DFIs is nD.

We calculate shares s_{i0}^{d} in the first five periods of our samples, setting aside the remaining 15 periods for estimation.

One way to understand what the supply-push instrument aims to achieve comes from writing actual DFI investment received by country i in period t as the product of the shares of DFI budgets and total DFI investments: $dfi_{it} = \sum_{d=1}^{nD} s_{it}^{d} D_{dt}$. This variable is endogenous in fixed effects estimation of equation 10 because the shares s_{it}^{d} respond to unobserved shocks to the investment climate. The number of investable opportunities in a country varies for reasons not wholly captured by the variable pc. An increase in the number of investable opportunities will attract more private investment and—under the first selection mechanism—also increase DFI investments in that country. So s_{it}^{d} will rise in response to an improvement in the investment climate. The result is the spurious (positive) correlation reported earlier. By replacing s_{it}^{d} by s_{i0}^{d} the instrument fixes the shares a country receives of DFI budgets in an initial period, thus blocking out the endogenous response of s_{it}^{d} to the sources of variation in investment demand that are not captured by the econometrician.

In our DGP, the strength of the instrument really comes from the persistence of country types. For example, type 3 countries receive a higher share of DFI investment in the initial period and will receive a higher share of any increase in the global DFI budget. DFI-specific country preferences add another source of persistence, but our results show it is less important for instrument strength in our set-up. If we eliminate the differences across countries by setting μ (the mean of the distribution from which project characteristics are drawn) equal to 2 for all countries, then the instrument loses all power, there is a large upward bias on the estimated coefficient, and the variance of estimated coefficients across iterations is enormous (double digits). But because these regressions almost always fail an 'F-test' (a statistic researchers would look at to assess instrument strength) we might hope the estimates would be disregarded—and in any case, having no persistent differences in investment climates across countries is not a realistic case.

More interesting is what happens when the instrument looks very strong, so would be deemed useable by researchers. The results reported in Table 3 are generated from the DGP where DFIs have zero additionality (the true β is zero). We report median β over 1000 iterations, because the occasional extreme outlier pulls the mean around. The first column of Table 3 shows that there is some upward bias to the estimated coefficient on DFI investments, when the project characteristics variable is excluded from the regression. This upward bias occurs because the instrument does not satisfy the 'exclusion restriction' required to isolate exogenous variation. As explained earlier, the demand-led nature of the DFI business model implies that its budget is endogenous. If the countries that a DFI has stronger preferences for get a positive draw from the stochastic process that determines investment opportunities, this DFI will invest more in total. The individual DFI budgets D_{dt} are therefore a function of unobserved shocks to expected returns in the countries that these DFIs have a preference for. This creates a positive correlation between u_{it} and $dfiIV_{it}$ and hence an inconsistent IV estimator. To see that this is the source of the bias, column 2 shows that, when we eliminate DFI-specific preferences for certain countries by setting $\phi = 0$, the median bias all but disappears.[27] In contrast, if we widen the differences in preferences between DFIs by increasing ϕ to 2, a larger bias results.

Staying with the case of $\phi = 2$, columns 4 and 5 illustrate what happens when the project characteristics variable is included in the regressions, with differing levels of noise. In column 4, the inclusion of this variable eliminates most of the bias, as was the case before. In column 5, where the variable is noisier, the bias reduction is smaller, and the null of zero additionality is rejected at a 5% significance level in more than 16% of replications.

[27] A similar result is obtained when we have only a single DFI in the DGP.

In the final three columns, we show that even with the project characteristics variable included, the remaining bias could be substantial. To illustrate this result, we reduce the time-series variation in the global DFI budget by halving σ_{db}, keeping $\phi = 2$. This change reduces the extent to which individual DFI budgets are driven by exogenous variation in the global DFI budget, thereby giving relatively more prominence to the endogenous response of DFI budgets to investment climate shocks in their preferred countries. It also reduces instrument strength, though the instrument is still far from weak: in column 6, for instance, the median F statistic is over 30 and the F statistic exceeds 10 in more than 95% of replications.

In column 6, where we do not control for project characteristics, this reduction in σ_{db} results in much greater bias. Now, even after controlling for the project characteristics variable (columns 7 and 8), a fair amount of bias remains, especially with the larger degree of noise in the project characteristics variable (column 8). In the latter case, the null of zero additionality is rejected at a 5% significance level in more than a quarter of replications. This bias is not due to instruments being weak, so it would not be uncovered by weak instrument diagnostic tests a researcher would routinely carry out as part of instrumental variable estimation.

Table 3: IV estimates: true $\beta = 0$

PC noise	Excl	Excl	Excl	Low	High	Excl	Low	High
σ_{db}	$0.2DB_1$	$0.2DB_1$	$0.2DB_1$	$0.2DB_1$	$0.2DB_1$	$0.1DB_1$	$0.1DB_1$	$0.1DB_1$
ϕ	1	0	2	2	2	2	2	2
Median β	0.11	0.01	0.20	0.05	0.11	0.56	0.17	0.32
Std Dev	0.43	0.48	0.41	0.21	0.31	0.61	0.33	0.47
% reject H0	12	6	22	12	16	41	20	29
% F>10	99	97	100	100	100	95	96	96
Median F	72	65	77	108	92	33	43	38

This table reports the median coefficients on DFI investments from IV regressions, together with their standard deviation, across 1000 replications, and the percentage that reject the null of zero additionality. The first row indicates whether there is measurement error added to the variable 'project characteristics' and the second row is the standard deviation of the random walk that determines DFIs' budgets and the third is the weight on country preferences. The bottom two rows give median F statistic and the percentage that are above 10, a common heuristic for instrument strength.

We caution against reading too much into the magnitudes of the biases we find. We have demonstrated that there are plausible circumstances in which there is a material upwards bias that might lead researchers to infer additionality when none exists (or a greater degree of additionality than exists) but we cannot say which configuration of our data generating process is the closest approximation to reality.

All the methods that we considered in this section to estimate the degree of DFI additionality using cross-country data are clearly open to the possibility of bias. The supply-push IV estimator appears to offer the greatest potential for bias reduction relative to naive OLS and FE estimators, but even this estimator is not risk-free: in the absence of good control variables, and if DFI budgets respond strongly to changes in the investment climate in a small number of 'preferred' countries, this estimator as well could yield misleading results. Hence, an application of this IV estimator will require careful consideration of the issues highlighted by our simulations.[28] In addition, our results suggest it could be fruitful to find natural experiments that allow researchers to identify exogenous variation in DFI budgets. Replacing DFI budgets in the construction of the supply-push instrument by an exogenous predictor for the DFI budgets that is not influenced by recipient country circumstances, should yield a consistent IV estimator. The recent turn towards the private sector in many donor governments' development finance strategies has resulted in many DFIs being given additional capital by their shareholders.[29] These recapitalisations could potentially be used for this purpose.

7 Firm-Level Evidence

Another place to look for evidence of additionality could be firm-level data. DFIs collect some data from their investee companies, and some countries run annual firm surveys that could be combined with these data; development agencies and DFIs may also commission researchers to assemble new firm-level datasets. In this section we argue that comparing levels of investment between DFI investees and 'control' firms would be fundamentally

[28] Temple and Van de Sijpe (2017) discuss how fragmentation on both the recipient and donor side can alleviate concerns about donor budgets responding too strongly to recipient country circumstances; it would be useful to undertake a similar analysis in this context before applying the IV estimator. Roughly speaking, if DFIs spread their investment budgets relatively evenly over a large number of countries, then it will be less likely that many DFI budgets are unduly influenced by the changing circumstances in just a handful of those countries. Similarly, if instrument values for each country depend on a large number of DFIs and these DFIs' budgets do not all react to shocks to expected returns in the country in the same way, then the correlation between the instrument and the error term will be weakened.

[29] In April 2018 World Bank shareholders announced they would inject \$5.5bn into the IFC; in October 2017 the UK government said it would give CDC Group an average of £620m-£703m annually over the next five years.

the wrong way to approach the problem and that trying to estimate the likelihood that a private firm would have made the same investment as a DFI is the more promising approach. Unfortunately, we will show that approach also produces a misleading answer, for two reasons.

Firstly, there is (again) the difference between what the econometrician can observe and the private information that investors use to guide their decisions. In some cases insiders will judge investment opportunities to be more appealing than they appear from the outside (in the language of our DGP: $er > pc$), and in other cases the opposite. Suppose DFIs are additional and only invest in projects below private expected return thresholds: around that threshold, on average DFIs are more likely to take the projects that are less appealing than they look, whilst the private sector is more likely to take those that are more appealing. Secondly, the econometrician is only likely to be able to gather data on firms that receive an investment from either a DFI or privately—firms that went looking for funding but failed to obtain it will be much harder to observe.

7.1 Two forms of Investment Data

Researchers may obtain data that record the external injections of funds into an enterprise, such as a new loan agreement or equity investment, perhaps through various private databases that track 'deals' such as that available from EMPEA (the Emerging Markets Private Equity Association) or Prequin, an independent data vendor. This is a natural fit with the nature of DFI investments, but has the drawback that it may be hard to distinguish between primary fundraising to finance new productive activity, and secondary investments, where existing shareholders are bought out or an existing loan is refinanced without any new funds being made available to the enterprise.

The second option is to gather investment data from firms' annual profit and loss (P&L) accounts, and balance sheets. These data would combine investment financed by internal cash flows with investment financed by external fundraising. When a firm raises money from DFIs, this will show up on its balance sheet at the time of investment, and then in the P&L over time, as it is spent on investment goods. Expenditure accounted for as investment may exclude some costs that could be considered as investment in the broader sense, such as funding operating losses during a start-up phase (working capital). If the object of interest is primary fundraising for new economic activity, investment expenditure in the P&L will not separate this from investment associated with the continuation of existing activities. Whether the researcher works with balance sheet or P&L data would depend on availability and choice of research question. A researcher could assemble a firm-level dataset with an indicator for when a firm receives an investment from a DFI, in an attempt to estimate the degree of additionality of DFI investment.

The data generating process we use in this paper is designed to capture discrete internal injections into enterprises, not how those funds are spent by firms over time.

7.1.1 The Level of Investment is the Wrong Place to Look

Evidence of additionality will not be found by comparing the levels of investment in firms that DFIs finance against 'control' firms that do not receive a DFI investment. The premise behind the idea of additionality is that there are some private enterprises that do not require public support, because they have their financing needs met by the private sector. In some cases the additionality of DFIs consists in their willingness to make a larger investment in a given project than private financiers are willing to, but in general there is no reason to expect firms that can raise private finance to invest any more (or less) than DFI investees. The quantity of each investment is largely determined by the needs of the underlying enterprise — whether those needs are met by a DFI or by private financiers has to do with expected returns and the competitive process between DFIs and the private sector.

If a regression was run and revealed that DFI investees on average do invest more than comparable firms that are financed privately, what would we learn about additionality? An estimate of how much more investment firms with DFI backers make, in comparison to firms with private backers, does not tell us whether those DFIs displaced private investors. It might be the case that DFIs have crowded out private investors, but because they are more patient investors their investees flourish and, as a result, tend to invest more. Statistical comparisons of DFI investees against control firms may be of interest for various reasons, but to identify pure investment additionality we need to estimate what would have happened had the DFI not invested. Firms that have either raised funds from elsewhere, or have not raised any funds at all because they had no opportunities that required funding, do not allow us to estimate that counterfactual.

7.1.2 We Would like to Know Expected Returns

Econometricians typically seek evidence of 'treatment effects'—what is the effect of a painkiller on a headache sufferer? But in this context, we do not want to know the effect of being 'treated' by a DFI; we want to know whether a private investor would have 'treated' a firm had the DFI not. The probability of receiving a 'treatment' has to do with an investment's expected (risk-adjusted) return. If researchers could observe these, and knew private investors' decision criteria, then we would know whether a private investor would have done that deal. Private investment decision criteria could be estimated by observing the projects that private investors have invested in. Can observable firm characteristics get us close enough to that?

7.2 What Could We Infer from Great Data?

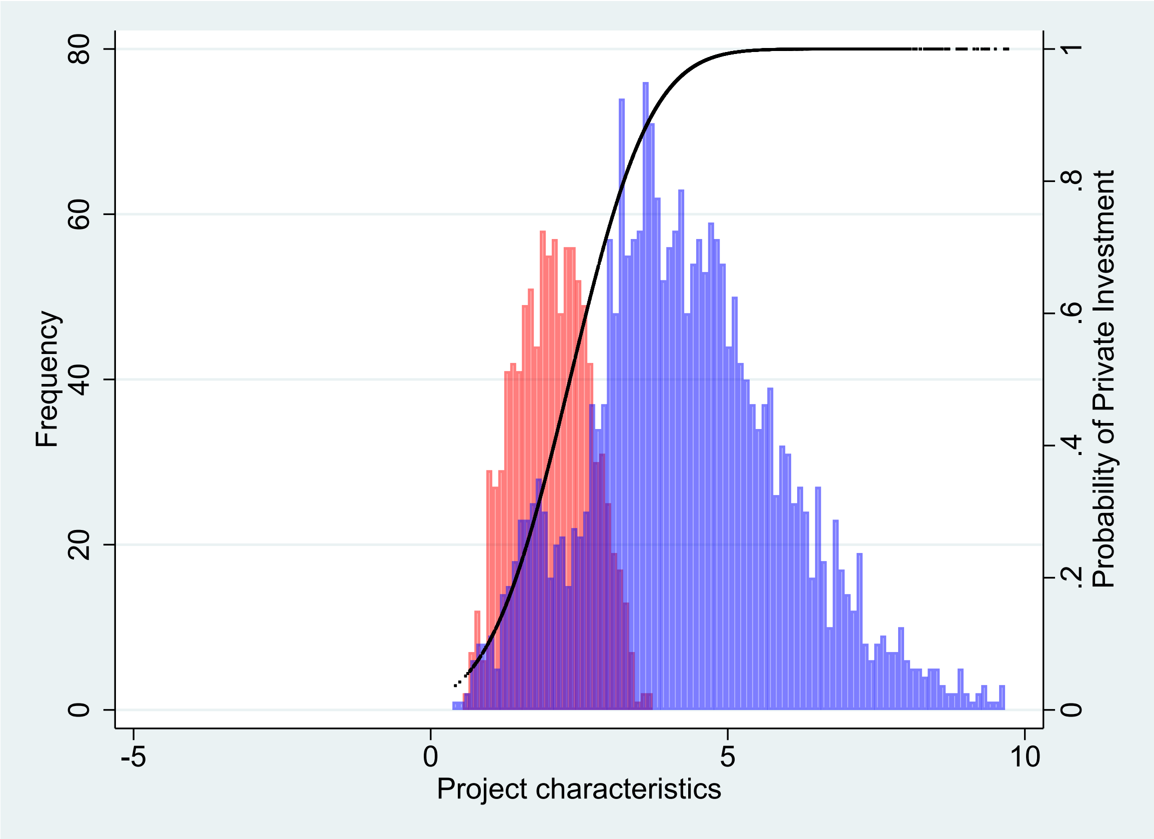
Suppose a researcher is in the enviable position of having access to a wonderful dataset: the record of every investment made by the DFI and every investment made by the private sector, as well as enough project-specific information to produce an unbiased estimate of the risk-adjusted expected return for every single project (our 'project characteristics' variable).

A natural place to begin is to estimate a discrete choice model where the funder's identity (DFI or private sector) is a function of the project characteristics. This will reveal any systematic differences in the types of projects supported by each funder. With this information, we can compute the predicted probability that a project with particular characteristics will receive one type of funding or the other. Specifically, for projects with characteristics like the DFI-funded projects, we can compute the predicted probabilities of being private sector-funded. If the probability is high, we might infer that the degree of additionality is low, and vice versa.

A researcher estimating a discrete choice model using our simulated data would recover a large positive and statistically significant coefficient on the observable project characteristics variable. These estimated coefficients are hard to interpret, and are easier to digest visually.

Figure 1 plots the predicted probabilities of 'treatment' by the private sector, estimated by a simple Probit model.[30]

Figure 1: Predicted probability of private investment



The distribution of estimated private returns (project characteristics) of DFI-funded projects is shown in red, and the distribution for privately funded projects is shown in blue. The black dots show the predicted probability that each project receives private sector funding, as read on the right-hand y-axis.

This plot looks exactly as the DFI would hope. The bulk of firms invested in by DFIs have 'project characteristics' in the range 1-3, and the predicted probability of private sector investment for such firms is quite low. The predicted probability that the private sector would have undertaken an investment climbs to 0.5 when the project characteristics reach around 2.5, the top of the DFI's range. The researcher could conclude that most DFI investments are probably additional, although there is a good chance (> 0.5) that roughly a third of the investments, with project characteristics towards the higher end of the range, crowded-out private investors.

[30] In this instance, the estimated coefficient on the project characteristics variable is 1.47 with a standard error of 0.07. The data are from our baseline parameterisations of the DGP, with firms pooled across 10 countries from a single period. We also increased the proportion of countries that are type 3, and set the DFI budget parameter to 0.2.

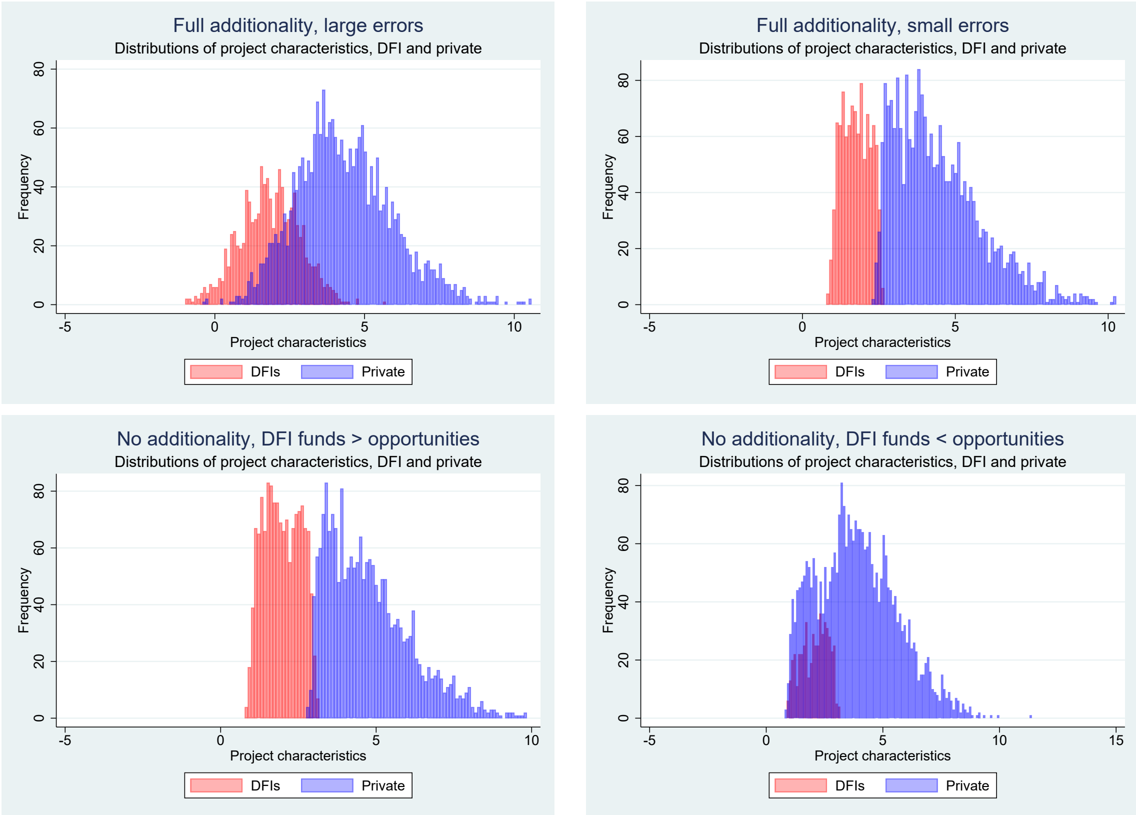
This conclusion would be incorrect, however. Although nothing prevents us from producing these predicted probabilities, to interpret them as evidence of additionality is to conflate what private investors did in the presence of DFIs with what they would have done in their absence. To make it clear why these two probabilities are fundamentally distinct, it is helpful to spell out the different mental models one might have about how investments are made.

Suppose private investors had been presented with all of the projects in our data set and then rejected some that were subsequently funded by the DFI. In that model, the fact that a project is DFI-funded is indeed revealing about its additionality: it is additional by construction. If, on the other hand, the DFI selects its investments first, then the DFI-funded projects are removed from the private investor's choice set. But if the DFIs can take projects before private investors get to make a decision about whether or not to fund them, the data cannot possibly reveal what the private investors would have decided. A hypothetical private investor that would have funded every DFI-funded project, or one that would have funded none of them, would both be equally consistent with the data.

As you may have anticipated, these encouraging results and predicted probabilities shown in figure 1 were estimated from the data generating process where DFIs have zero additionality: the probability that the private sector would have invested, had the DFI not, was 1 for all projects in the data. Figure 2 further illustrates this point with four possible distributions of project characteristics for private and public investments. Both plots in the top row are from a data generating process with full additionality; in the bottom row the data generating process has zero additionality. The two right-hand panels show cases where reality and appearance coincide. The two left-hand panels look almost the same, but the underlying truth is reversed. The fact that two completely different additionality scenarios can give rise to virtually identical patterns in the data clearly tells us that observable project characteristics are a poor guide to additionality.

This leads us to the discomforting conclusion that even possessing fantastically rich and accurate firm-level data, obtaining rigorous evidence of additionality might be very difficult. The good news is that we can at least imagine a data set that might be more informative about additionality. In addition to our already comprehensive data set, we

Figure 2: Any distribution of private and DFI projects is possible



The top row shows the distribution of project characteristics for private and DFI investments when DFIs are fully additional, the bottom row when they have zero additionality. In the top row: on the left the 'errors' between observed project characteristics and the privately-observed expected returns are large, on the right small. In the bottom row: on the left the DFI budget exhausts all projects in its eligible set, on the right the DFIs' budget can only cover a small share of the investments they would like to make.

would also need DFIs to provide information about all the projects that they considered but rejected. Observing which of these opportunities private investors subsequently invest in, or reject, would allow us to make valid statements about the likelihood that certain types of projects will receive private funding, in the absence of DFI investment. To illustrate how, suppose, for instance, that the DFI funds half of all firms with project characteristics of 2.5 and the private sector subsequently funds half of those rejected by the DFI. We do not know whether the DFI rejected those opportunities because it observed that true expected returns were too high, and it wished to avoid crowding out the private sector, or to the contrary, because expected returns were too low. Our best estimate would be that the private sector invests in firms with project characteristics of 2.5 with a probability of 0.5, but the true probability could be as low as 0.25 (if the private sector would have rejected all of those that the DFI funded) or as high as 0.75 (if the private sector would have funded all of those that the DFI did). Before we knew only that the probability could be anywhere from zero to one, but now we can tighten the bounds and come up with a plausible central estimate. This calculation is possible because we can see the full set of opportunities the private sector had the option to invest in, not just the ones that actually received their investment.

The analysis in this section may appear abstract, but its message is concrete: caution is needed when comparing DFI investments against private comparators because differences do not necessarily reveal additionality and similarities do not necessarily reveal crowding-out. DFIs' investments may look different from private, but that could be because they are taking all available deals of a certain type, even though private investors would have done them. DFIs' investments may resemble private investments, yet still be additional for any number of reasons to do with the difference between what insiders based their decisions on, and superficial observable characteristics such as country, sector and investment type. For example, private investors may support the first entrant into a market but not a second, because the first has established advantages that reduce new entrants' expected returns—yet a DFI may be willing to support a second entrant because creating competition in markets is a strategic objective.

We have shown that when judged by observable project characteristics in certain circumstances DFIs may appear to be additional but not be, or appear not to be when they are. We have not shown these circumstances are all equally likely. One might still reasonably believe that investments that look unlike deals that private investors are doing are more likely to be additional than deals which look the same. The burden of proof may be higher when DFIs are active in markets where private investors are active.

Thus far we have sought to take a rigorous approach to identifying the causal impact of DFI investments, but if such evidence is unobtainable then we must switch mindset, and consider circumstantial evidence that might not prove anything, but which might shift the subjective probability in the mind of a rational individual.

8 Qualitative Evidence

In practice, available evidence of additionality is qualitative.[31] The primary source of evidence is the quality of the arguments made that additionality is present at the time that the decision to invest is made, although DFIs do also conduct retrospective surveys of their investees and private investors that ask questions about additionality.

The definitions employed in evaluations of DFIs' additionality are broad and encompass both investment (quantity) additionality and development (quality) additionality. For example the world's largest DFI, the World Bank's IFC, defines additionality as the "benefit or value addition we bring that a client would not otherwise have. In other words, our additionality is a subset of our role that is unique to IFC and that cannot be filled by the client or any commercial financier."[32] A technical report for the European Union defines additionality as "in its most basic sense ... the net impact of an intervention after taking into account what would have happened in the absence of the intervention (reference case)." DFIs wishing to deploy EU grants are asked to 'demonstrate the additionality' by 'providing evidence' but in practice that consists of completing a questionnaire in which the DFI explains what the grant will achieve, which may include the assertion that private financiers would not have backed the project (EUBEC, 2013). When evidence for the additionality of EU 'blending' is evaluated, it consists of assessing the quality of those arguments (ADE, 2016).

As already mentioned, one difference between how DFIs approach additionality in practice and the strict interpretation of investment (quantity) additionality used in this paper is that many DFIs look at what they call 'financial' additionality, namely whether they are providing a form of finance that the market would not. This is necessary for investment additionality—if a DFI invests in a project that private investors would not, then by definition it is providing finance on terms the market is not—but it is not sufficient: a DFI may crowd-out private investors by offering finance on more favorable terms than private investors would.[33] For example, DFIs may observe that they offer longer-term loans than

[31] See Spratt and Collins (2012), Independent Evaluation Group (2008) and ADE (2016). We are aware of no attempt to identify additionality in investment data.

[32] IFC (2013).

[33] In theory a project sponsor might reject an offer from a private financier then accept the same financial offer from a DFI, because the DFI brings non-financial benefits, in which case financial additionality is not

anything available in the local market, but if project sponsors prefer long-term finance they may choose DFIs even if they would have made do with short-term debt from a private source, in the absence of a DFI. Strictly speaking, investment additionality would require the accompanying judgement that the project would not have been viable in the absence of long-term finance.

Some development banks ensure that their debt is priced above market comparators, because that proves they are providing some additional value (otherwise the project sponsor would not agree to pay more). But because the sum total of what they offer—the loan plus supplementary benefits such as de facto political risk mitigation—is evidently more appealing to project sponsors than what private investors would offer, the DFIs may be crowding-out private finance and fail to have investment additionality. Pricing above market is evidence of 'development' additionality—for example the project may be more likely to succeed with DFI support, because of the supplementary benefits that they confer.

In some respects, the best evidence of additionality consists of the tacit knowledge of trusted market participants who are familiar with investor behaviour and can say whether, or not, private investors would have made a given investment. Although investment professionals within DFIs with familiarity with their markets may be able to spot claims of additionality that are obviously not credible, when fine-grained judgements based on intimate familiarity with the project in question are required, a problem for DFIs is that only the transaction team involved will possess such knowledge, yet these are the individuals that can have an incentive to claim additionality is present when it is not.

Surveys of project sponsors and competing private investors may be informative, but cannot be entirely relied upon. Although some project sponsors may start knowing that only a DFI would support them, or come to learn that, most entrepreneurs are almost by definition people who think that they have a fabulously exciting investment opportunity that people will want to invest in. Their subjective perceptions of their own appeal cannot be taken at face value. On the other hand, if survey respondents realise that 'this project would not have gone ahead without DFI support' is the answer that evaluators are looking for, they may give that answer in an effort to please. Private investors may be equally unreliable—if given examples of successful DFI investments they may be prone to claiming

strictly necessary. We regard that slim possibility as the exception that proves the rule.

they would have made what looks like the right decision (of investing) in retrospect.

A method known as 'process tracing' has potential in this context. Informally, the idea is to specify in advance what one would expect to observe if a hypothesis is true, and what would tell you that it is probably not true, and then to assign different weights to these factors and examine individual cases in that light. This method can generate numerical estimates of the probability that a hypothesis is true.[34] One challenge for using this method to assess additionality could be that the investment process has relatively few moving parts and there simply might not be enough observable factors to reach much of a conclusion one way or another.

9 How to Think About Additionality

Everyone wants evidence that aid is well spent and that development interventions achieve their objectives. The 'credibility revolution' in econometrics has put the identification of causal impacts at the forefront of the field. These are good things. Where evidence can be obtained (at a reasonable cost) it certainly should be. But there are some questions that research cannot answer. Governments want 'evidence-based policy-making' but—we would argue—the reality is that more often than not, policy will be made in the presence of considerable uncertainty. Facing this fact could improve how policy is made: see Manski (2013) for some ideas of how.

Obtaining evidence of additionality is not only a problem for DFIs (although they are more frequently criticised for the lack of it).[35] In truth the additionality of traditional aid is also sometimes uncertain, especially over the long run: a donor funding a health clinic cannot be sure the government would not have done so in its place.[36] We cannot even be entirely certain that aid is beneficial overall, on average across countries—although the weight of the evidence points in that direction, it leaves room for doubt.[37] So the fact

[34]See Collier (2011) for an introduction; Humphreys and Jacobs (2015) and Befani and Stedman-Bryce (2017) for more quantitative Bayesian applications.

[35]See for example 'Civil society groups set out key recommendations ahead of OECD discussion on changes to aid rules' on the website of Eurodad, a network of civil society organizations, which asks that evidence of additionality is provided before private sector instruments can be counted as aid. The same is not so often asked of traditional aid.

[36] In the aid literature, the possibility that aid intended for one purpose may be used for another is known as fungibility: see Van de Sijpe (2013).

[37] See Temple (2010) for a survey of the evidence. Eubank (2012) provides the striking counterexample of

that definitive evidence of additionality may always prove elusive should not alarm the development community, where uncertainty is often the norm.[38]

Given this, we should stop talking about 'measuring' additionality, and start asking under what circumstances we think additionality is more likely. Acknowledging uncertainty may not be easy for DFIs, most of whom regard additionality as a binary and only authorise investments when they are confident of it.[39] Boards of directors and shareholders may not be ready to hear: 'we think there is 50 per cent chance this project is additional'. But even if additionality is a binary, one may form a subjective estimate of its probability. In reality investment committees will inevitably be more confident in some cases than in others.

If investment additionality was treated as a probability, then it could be traded-off against other objectives, such as development additionality. Investment committees could take decisions on the basis of expected development impact. For the sake of illustration, suppose a DFI confers development additionality Δ whether investment additionality is present or not, and other development outcomes Ω that are conditional on investment additionality. Development impacts Ω that are conditional on investment additionality are those that would still be there if the private sector had taken the project, such as creating jobs, paying taxes and producing goods and services. If the private sector would have delivered them, DFIs achieve nothing by doing so in their place. Development impacts Δ are those that reflect how DFIs do things differently to private investors, such as having higher environmental and social standards, creating higher quality jobs, and influencing the production of goods and services in a way that does more to alleviate poverty.

If the probability of investment additionality is π_a then the expected development impact is $\pi_a(\Delta + \Omega) + (1 - \pi_a)\Delta$. So sometimes uncertainty about additionality could be compensated for by larger expected impacts conditional on investment additionality, or great development additionality in the absence of it. Of course putting numbers on these things can be specious, but it may be possible to admit varying degrees of confidence about

Somaliland, whose unrecognised status meant it was denied aid, and which went on to flourish in comparison to most aid recipients.

[38] Of course even when the correct model is estimated on sample of data, there is uncertainty around the estimated parameter—that's what confidence intervals are for. Here we mean uncertainty in a more radical sense.

[39] We are aware of three DFIs that rate ex-ante expected additionality on a scale.

investment additionality into decision-making, in a pragmatic way. The challenge will be to define sufficiently objective criteria for placing rough magnitudes on probabilities and impacts.

In the absence of proof beyond reasonable doubt, we should seek circumstantial evidence that would cause a reasonable person to update their (probabilistic) belief that additionality is present. The short discussion of qualitative evidence in the proceeding section found fault with everything, but imperfect evidence is better than nothing. Surveys convey some information. If we can assume that project sponsors' 'optimism bias' is some unknown constant then even if percentages of respondents who answer questions one way or another cannot be trusted, variation across time and place can still be informative. Financial additionality is not evidence of investment additionality, but it is better to have information about transactions in the markets where DFIs are active, than not.

9.0.1 Good Enough Evidence

Anyone looking for evidence that DFIs are additional should be satisfied with having reason to believe they are 'merely' likely to be, for two reasons. The first reason has been the focus of this paper: definitive proof is impossible to obtain. The second is that we do not want DFIs only to invest when they are certain of additionality. Suppose DFIs knew when a project had only a 50% chance of being additional. If they turn away 10 projects like that a year, that's 5 projects that would have helped accelerate investment in poor capital-scarce countries, that did not happen.[40] Traditional aid projects should also tolerate risks, of waste, inefficiency and corruption, for similar reasons. Quite what the threshold of probability ought to be is hard to say, but it is not 1. In practice the decision to invest would also incorporate development impacts that are not conditional on investment (quantity) additionality, as suggested above.

[40] And because DFIs target positive financial returns, these risks are decidedly asymmetrical. Assuming they do no harm, when DFIs do crowd-out private investors then their shareholder governments still make a profit and can recycle their money. That's not the goal, but it's not a cost in the same sense as wasting grant funding.

9.0.2 Where DFIs can Improve

There are two main areas for improvement: DFIs could do more to collect and share data on what sorts of deals private investors are doing in the markets in which they operate; and when DFIs are doing deals that resemble those that private investors are doing, they should more carefully document their reasons for thinking they have a reasonable chance of being additional, in a format that can be made public.

10 Conclusion

This paper has shown why definitive evidence of additionality will very likely remain elusive. We have concerned ourselves only with investment (quantity) additionality and articulated a data generating process which we believe captures the important aspects of DFIs' demand-led business model. Using these data as an artificial laboratory, we have run quantitative 'experiments' to test how well standard estimating techniques recover additionality.

In a cross-country setting, estimators that make no attempt to control for endogeneity can report large and statistically significant positive or negative estimated coefficients, whether additionality is present or not, depending on the procedure that DFIs use to select their investments. These biases are especially large when the researcher only has access to control variables that imperfectly predict the variation in private investment across countries. We have tested two common techniques for purging endogeneity—a GMM estimator that uses internal instruments and a 'supply-push' IV estimator. Both of these prove to be unreliable, although the supply-push IV estimator appears to offer the greatest potential for bias reduction. The supply-push estimator is biased because the demand-led nature of DFIs' business models means that their budgets are not entirely exogenous.

It is also impossible to reliably infer additionality from firm-level data. We tested the approach of estimating the probability that private investors would have done a deal based on observable project characteristics, and showed how this produces misleading results. The considerations that investors base their decisions on cannot be deduced from the information about projects that a researcher could hope to obtain, because DFIs may invest in all projects of a certain type, even if private investors would have taken those deals.

We argue that qualitative data is also unreliable. Surveys that ask private investors and entrepreneurs what would have happened in the absence of DFI investments cannot be taken at face value.

DFIs making investment decisions and those looking for evidence of additionality should embrace uncertainty, because it is inevitable. The Sustainable Development Goals and accompanying Financing for Development strategy imply that we do not want DFIs making investments only when they are certain of additionality, because that would entail making the 'type 1' error of turning away opportunities to accelerate investment in capital-scarce developing countries. The need for investment is so great that we cannot afford missed opportunities.

The idea of estimating additionality from observational data is not entirely hopeless, and careful application of IV estimators may sway our beliefs just like any other instance of not wholly definitive evidence. But for practical purposes, because additionality cannot be observed or 'measured' we should instead look for circumstantial evidence that DFIs are investing in a way that would lead us to believe additionality is likely to be present. Our analysis of firm-level evidence showed how comparing investments made by DFIs and private investors can mislead, but nonetheless DFIs should try to provide information about how the investments that they make compare to those made by private investors, in the markets where they are active, and when they make investments that superficially resemble those made by private investors DFIs should document the reasons why they believe they are likely to be additional.

Because investment professionals within DFIs are best placed to judge whether an investment is additional, the incentives within DFIs matter. If DFIs are under pressure to hit volume targets, or staff incentives are tied to volumes, it will be harder to hold the line on additionality than if DFIs are mission oriented and staff only wish to deploy capital where they believe it will have development impact.

References

ADE (2016). Evaluation of blending. European Commission, ADE S.A. on behalf of the Evaluation Unit of the Directorate-General for International Cooperation and Development.

Angrist, J. D. and J.-S. Pischke (2010). The credibility revolution in empirical economics: how better research design is taking the con out of econometrics. *Journal of Economic Perspectives 24*(2), 3–30.

Arellano, M. and S. Bond (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *Review of Economic Studies 58*(2), 277–297.

Arellano, M. and O. Bover (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics 68*(1), 29–51.

Athey, S. and G. W. Imbens (2017). The state of applied econometrics: Causality and policy evaluation. *Journal of Economic Perspectives 31*(2), 3–32.

Bartik, T. J. (1991). *Who benefits from state and local economic development policies*. Kalamazoo, Michigan: W.E. Upjohn Institute for Employment Research.

Bazzi, S. and M. A. Clemens (2013). Blunt instruments: Avoiding common pitfalls in identifying the causes of economic growth. *American Economic Journal: Macroeconomics 5*(2), 152–86.

Befani, B. and G. Stedman-Bryce (2017). Process tracing and bayesian updating for impact evaluation. *Evaluation 23*(1), 42–60.

Blundell, R. and S. Bond (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics 87*(1), 115–143.

Bond, S. R. (2002). Dynamic panel data models: a guide to micro data methods and practice. *Portuguese Economic Journal 1*(2), 141–162.

Bun, M. J. and F. Windmeijer (2010). The weak instrument problem of the system gmm estimator in dynamic panel data models. *Econometrics Journal 13*(1), 95–126.

Collier, D. (2011). Understanding process tracing. *PS: Political Science & Politics 44*(4), 823–830.

Deaton, A. and N. Cartwright (2017). Understanding and misunderstanding randomized controlled trials. *Social Science & Medicine*.

Dreher, A. and S. Langlotz (2015). Aid and growth. new evidence from an excludable instrument. Centre for Economic Policy Research (CEPR), Discussion Paper No. 10811.

Eubank, N. (2012). Taxation, political accountability and foreign aid: Lessons from Somaliland. *Journal of Development Studies 48*(4), 465–480.

EUBEC (2013, March). Enhancement of blending activities initial topic: measuring results, monitoring and reporting. Technical report, Report by technical group 2 (TG2) to the policy group of the European Union blending and external cooperation platform (EUBEC platform).

Hainz, C. and H. Hakenes (2012). The politician and his bankerâĂŤhow to efficiently grant state aid. *Journal of Public Economics 96*(1-2), 218–225.

Holtz-Eakin, D., W. Newey, and H. S. Rosen (1988). Estimating vector autoregressions with panel data. *Econometrica 56*(6), 1371–1395.

Humphreys, M. and A. M. Jacobs (2015). Mixing methods: A bayesian approach. *American Political Science Review 109*(4), 653–673.

IFC (2013). Additionality primer. IFC Memo.

Imbens, G. (2018). Comments on: "understanding and misunderstanding randomized control trials" by Deaton and Cartwright. Standford University Mimeo.

Independent Evaluation Group (2008). Independent evaluation of IFC's development results 2008: IFC's additionality in supporting private sector development. World Bank Group.

Kenny, C., J. Kalow, J. Leo, and V. Ramachandran (2018). Comparing five bilateral development finance institutions and the IFC. Center for Global Development Policy Paper.

Kraay, A. (2015). Weak instruments in growth regressions: Implications for recent cross-country evidence on inequality and growth. World Bank Policy Research Working Paper 7494.

Manski, C. F. (2013). *Public policy in an uncertain world: analysis and decisions*. Harvard University Press.

Massa, I., M. Mendez-Parra, and D. W. te Velde (2011). The role of development finance institutions in tackling global challenges. ODI Discussion Paper.

Massa, I., M. Mendez-Parra, and D. W. te Velde (2016). The macroeconomic effects of development finance institutions in sub-Saharan Africa. ODI Discussion Paper.

McKenzie, D., N. Assaf, and A. P. Cusolito (2017). The additionality impact of a matching grant programme for small firms: experimental evidence from Yemen. *Journal of Development Effectiveness 9*(1), 1–14.

Nunn, N. and N. Qian (2014). U.S. food aid and civil conflict. *American Economic Review 104*(6), 1630–1666.

Rodrik, D. (2015). *Economics rules: The rights and wrongs of the dismal science*. WW Norton & Company.

Roodman, D. (2009a). How to do xtabond2: an introduction to difference and system gmm in stata. *Stata Journal 9*(1), 86–136.

Roodman, D. (2009b). A note on the theme of too many instruments. *Oxford Bulletin of Economics and Statistics 71*(1), 135–158.

Roodman, D. (2015). xtabond2: Stata module to extend xtabond dynamic panel data estimator.

Spratt, S. and L. R. Collins (2012). Development finance institutions and infrastructure. Private Infrastructure Development Group.

Temple, J. (2010). Aid and conditionality. In *Handbook of development economics*, Volume 5, pp. 4415–4523. Elsevier.

Temple, J. and N. Van de Sijpe (2017). Foreign aid and domestic absorption. *Journal of International Economics 108*, 431–443.

Van de Sijpe, N. (2013). Is foreign aid fungible? evidence from the education and health sectors. *World Bank Economic Review 27*(2), 320–356.

Warner, A. (2013). A framework for efficient government investment. IMF Working Paper WP/13/58.

Werker, E., F. Z. Ahmed, and C. Cohen (2009). How is foreign aid spent? Evidence from a natural experiment. *American Economic Journal: Macroeconomics 1*(2), 225–244.

Windmeijer, F. (2018). Testing over- and underidentification in linear models, with applications to dynamic panel data and asset-pricing models. Bristol Economics Discussion Papers 18/696, Department of Economics, University of Bristol, UK.