The University Of Sheffield.

Department Of Economics.

Sheffield Economic Research Paper Series.

The Elusive Quest for Additionality

Authors: Patrick Carter, Nicolas Van de Sijpe, Raphael Calel

ISSN 1749-8368

SERPS no. 2019022

December 2019

The Elusive Quest for Additionality*

Patrick Carter[†] Nicolas Van de Sijpe[‡] Raphael Calel[§]

13th December 2019

Abstract

Development finance institutions (DFIs) annually invest $90 billion to support under-financed projects across the world. Although these government-backed institutions are often asked to show that their investments are "additional" to what private investors would have financed, it is rarely clear what evidence is needed to answer this request. This paper demonstrates, through a series of simulations, that the nature of DFIs' operations creates systematic biases in how a range of estimators assess additionality. Recognising that rigorous quantitative evidence of additionality may continue to elude us, we discuss the value of qualitative evidence, and propose a probabilistic approach to evaluating additionality.

JEL codes: F21, F35, G15, O16, O19.

Keywords: Development Finance Institutions; Investment; Additionality; Simulation.

*For helpful comments and advice, we thank Colin Buckley, Markus Eberhardt, Georgios Efthyvoulou, Rune Jansen Hagen, Arne Risa Hole, Alex MacGillivray, Vijaya Ramachandran, Laurence Roope, Jon Temple, Frank Windmeijer, Adrian Wood, an anonymous referee, and numerous seminar participants. Part of the work on this paper was completed whilst Paddy Carter was a Research Fellow at the Center for Global Development working on a programme funded by the Bill and Melinda Gates Foundation. The other two authors consulted for the Center for Global Development on the paper. At the time of writing, Paddy Carter is Director of Research and Policy at CDC Group, the UK's Development Finance Institution. The views expressed in this paper are those of the authors alone.

[†]CDC Group. Email: pcarter@cdcgroup.com.

[‡]University of Sheffield, Department of Economics. Email: n.vandesijpe@sheffield.ac.uk.

[§]Georgetown University, McCourt School of Public Policy. Email: raphael.calel@georgetown.edu.

1 Introduction

Hopes of achieving the Sustainable Development Goals rest, in large part, on the international community's ability to direct unprecedented investment flows to poor, capital-scarce nations. Against the background of a long-standing debate about the effectiveness of foreign aid (see e.g. Bourguignon and Sundberg, 2007; Doucouliagos and Paldam, 2009), development finance institutions (DFIs) have emerged as critical players in a global effort to leverage public funds to bring trillions of dollars of private investment into developing countries.[1] DFIs annually make investments worth around $90 billion (Runde and Milner, 2019), but an important question remains unanswered: do DFIs increase total investment in developing countries, or do they merely displace private investment?

In this paper, we interrogate a number of methods that assess whether DFI investments are "additional" to what the private sector would have provided. DFIs are routinely asked to demonstrate the additionality of their investments, and are frequently criticised for being unable to do so (see, for instance, Countdown 2030 Europe, 2018; Griffiths et al., 2014; Pereira, 2015). What acceptable evidence would look like, or how it could be obtained, is, however, rarely articulated. We answer this question by formalising DFIs' investment decisions in a data-generating process (DGP), and then conducting simulations to evaluate a range of econometric strategies for assessing additionality. Using simulated datasets for which we know the true extent of additionality, we show how various estimators fail to recover the truth even with much better data than researchers typically have access to. The specific way in which DFIs operate alongside private investors almost automatically introduces bias into the assessment of DFIs' investment additionality, and this bias is not easily removed by using more sophisticated estimators.

Our simulations have three main parts. First, a list of projects is created, each project with some expected return on investment. Second, DFIs allocate their finite budget to projects located within some band of expected returns, while private investors finance projects with expected returns above a minimum threshold. Finally, a researcher observes the expected returns of individual projects with some error, and has to work out whether the projects financed by DFIs would otherwise have received private investment.

We first demonstrate that coefficients estimated by OLS and fixed effects in cross-country investment regressions provide a misleading measure of DFIs' additionality. We then show that more sophisticated econometric methods commonly used in the broader aid effectiveness literat-

[1] 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 African Development Bank et al. (2015).

ure, in particular instrumental variable (IV) estimation that relies on a supply-push instrument and system Generalised Methods of Moments (GMM) estimation, may also produce misleading results in this context. For the supply-push IV estimator, we establish that the exogeneity of the instrument is undermined when a DFI's total investment budget responds to shocks in countries that the DFI has a strong link with. The combination of changes in countries' investment environments and trends in the global DFI budget also spells trouble for this estimator. In system GMM, relatively persistent changes in countries' investment climates invalidate the moment conditions that the estimator relies on for consistency.

We further examine the possibility of identifying additionality in project-level data. We first consider and fail to find a basis for establishing additionality in comparisons of investment levels between firms funded by DFIs or private investors. We then test the idea of using observable project characteristics to estimate the probability that private investors would have undertaken DFI-funded projects. If this probability is low, we might be tempted to conclude DFIs are mostly additional. We show, however, that, when project characteristics are noisy measures of expected returns, DFI and private investments may superficially resemble each other, giving the impression of low additionality, even if DFIs are fully additional. Conversely, DFI and private investments may look very different if DFIs fund all projects of a certain type, even if DFIs are crowding out private investors. Our DGP is able to generate datasets that look indistinguishable to a researcher, even though the underlying true degree of additionality is completely different. Our assessment of the qualitative data is similarly skeptical. We argue that most self-reported evidence, both from DFIs and their investees, cannot be seen as definitive.

Our results suggest that unequivocal evidence of DFIs' investment additionality will remain elusive. In light of this, we propose that DFIs should take a probabilistic approach to evaluating additionality, focusing on identifying the circumstances in which an investment is more or less likely to be additional. Process tracing approaches could be helpful to structure the – often circumstantial – evidence and to translate it into a probability that an investment is additional. We discuss how this probability of additionality can then be evaluated together with other aspects of the investment project when making an investment decision, and how this simple step could improve DFIs' decision-making.

2 Background literature

2.1 How development finance institutions work

DFIs fulfill a number of roles and this paper is concerned with only one of them: primary fundraising intended to finance new economic activity, involving the installation of physical capital, investment in intangible capital, working capital to cover start-up losses, and so forth. DFIs sometimes provide secondary financing as well, by refinancing existing loans or providing an exit for earlier-stage investors, which implies a change of creditor or ownership without any new funds being raised for the enterprise itself. DFIs also seek 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 financing in markets where there is a shortage (see e.g. Dalberg Global Development Advisors, 2010; Griffith-Jones, 2016). Although these activities are all indirectly aimed at increasing the quantity of investment, for the sake of simplicity we focus on DFIs' primary, direct investment only.

DFIs have a demand-led investment model (Kenny, 2019*a*; Savoy, Carter and Lemma, 2016). They typically rely on others – known as project sponsors – to come up with investment ideas and come looking for money. The size of the investment 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).

The traditional DFI model is to invest on commercial terms (see e.g. Attridge and Engen, 2019; Kenny, 2019*a*; The Association of European Development Finance Institutions, 2016). Concessional finance is also available, but most DFIs draw a sharp distinction between concessional financing and their main business. 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.

DFIs usually have a mandate from shareholders to be self-financing (see e.g. Spratt and Collins, 2012; Xu, Ren and Wu, 2019). Many DFIs also operate on the premise that investing on commercial terms is good for development (The Association of European Development Finance Institutions, 2016). 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.

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.

Project sponsors may care about many things, but somewhere towards the top of the list is obtaining finance on the most favourable terms. DFIs' pricing is not always more favorable than private investors, but they also offer a range of non-financial benefits, such as political protection (Hainz and Kleimeier, 2012), which can make them more attractive. Not everything about DFIs is more appealing, though. DFIs impose higher environmental, social and governance standards (Kingombe, Massa and te Velde, 2011); they ask investees to report development outcomes, which is costly; and they may also interfere with corporate strategy.

2.2 Assessing additionality

The simplest definition of additionality is to make an investment happen that would not have happened in the absence of the DFI's intervention. The challenge in assessing additionality lies in establishing the counterfactual of what would have happened without the DFI's investment.

DFIs and multilateral development banks (MDBs) have themselves recently proposed a methodology for reporting the amount of private finance mobilised by their investment (African Development Bank et al., 2017; Multilateral Development Banks and European Development Finance Institutions, 2018), and the OECD's Development Assistance Committee is working on a separate international standard to measure the same (Benn, Sangaré and Hos, 2017).[2] To determine additionality, these approaches appear to rely chiefly on the type of financing provided by DFIs. For instance, for any syndicated loan (a loan offered by a group of lenders) led by a DFI, the MDB proposal counts all of the private contribution to the loan as mobilised private financing, effectively assuming that, in the absence of the loan, no private financing would have been offered. These methodologies are not so much exercises in assessing additionality, then, as in asserting it.

If we approach the question of additionality more agnostically, a natural starting point would be to look for its most obvious implication. If DFI investments are additional, we would expect them to increase the total amount of investment in developing countries, which could be assessed

[2] Against the background of the Copenhagen Accord's commitment by developed countries to mobilise $100 billion annually by 2020 to support developing countries' efforts to address climate change, there is also a related ongoing debate on the measurement of the amount of climate finance mobilised by developed country interventions (see e.g. Brown et al., 2015; Haščič et al., 2015; Jachnik, Caruso and Srivastava, 2015).

by estimating a version of the following equation:

$$I_{it} = \beta dfi_{it} + \gamma \widetilde{pc}_{it} + \delta_t + w_i + u_{it} \quad (1)$$

where I_{it} is total investment in country i in period t, dfi_{it} is the amount of DFI investment received, δ_t is a set of time dummies with associated coefficients, w_i is a time-invariant country effect, and u_{it} is the transient error. $\widetilde{pc}_{it}$ is a control for the observable characteristics of investment projects, to be discussed in more detail later. If DFI investment displaces private investment then $\beta < 1$, with $\beta = 0$ corresponding to zero additionality. Conversely, $\beta > 1$ if DFI investment catalyses private financing by helping to create active capital markets or by demonstrating to sceptical private investors where good returns can be made in developing countries.

Using a version of this model, te Velde (2011) estimates positive effects of DFI investment (as a share of recipient GDP) for some DFIs, but not for others. Massa, Mendez-Parra and te Velde (2016) also report positive and significant coefficients for a subset of DFIs, but no statistically significant effect when the investments of DFIs are pooled. Broccolini et al. (2019) estimate versions of equation (1) at the country-sector-year level, finding that the participation of a multilateral development bank in a syndicated loan increases the number of loans and the amount of syndicated lending (as a % of GDP) in subsequent years within the same country and sector. The causal claims in these papers mostly rest on the ability of their fixed effects to absorb all unobserved factors that correlate with both DFI investment and total investment.[3]

In our simulations, we will evaluate the ability of OLS and fixed effects estimation to recover the true degree of additionality in equation (1). We will also examine what we can learn from two estimation methods that are popular in the broader aid effectiveness literature and that can readily be applied to estimate the degree of additionality, namely supply-push instrumental variables (IV) and system Generalised Methods of Moments (GMM) estimation.

One approach to identify additionality is to look for an external instrument that is sufficiently strongly correlated with DFI investment but is uncorrelated with the error term and has no independent effect on the overall quantity of investment. Finding such an instrument in a cross-country context is notoriously difficult (see e.g. Bazzi and Clemens, 2013), but our setting has the advantage of offering a natural candidate in the form of a supply-push instrument. This instrument relies on the idea that the budgets of DFIs fluctuate for reasons that are unconnected to

[3]The sectoral dimension in Broccolini et al. (2019) allows the authors to estimate specifications that absorb all time-varying country-specific and time-varying sector-specific unobservables, as well as all time-invariant country-sector-specific unobservables. In Massa, Mendez-Parra and te Velde (2016), only time-invariant country-specific factors are absorbed, while te Velde (2011) relies on a random effects estimator that does not filter out w_i.

changes in the investment climate in recipient countries, and that DFIs have persistent preferences for some countries over others. When a DFI's overall investment budget increases, then, some recipient countries will experience a larger increase in DFI investment than others for reasons that should be uncorrelated with their domestic circumstances. Variants of this instrument have been used by several recent papers to estimate the macroeconomic effects of foreign aid (Dreher and Langlotz, 2017; Nunn and Qian, 2014; Temple and Van de Sijpe, 2017; Werker, Ahmed and Cohen, 2009). Spratt et al. (2019) also suggest it could be possible to use a supply-push instrument to estimate the macroeconomic effects of DFI investment (citing, in support, an earlier version of our paper that was more optimistic about the performance of this instrument).

In the literature on the effectiveness of foreign aid, difference and system GMM estimation are also commonly used to deal with potential endogeneity in cross-country panel regressions (for examples, see Dalgaard, Hansen and Tarp, 2004; Djankov, Montalvo and Reynal-Querol, 2008; Dreher, Nunnenkamp and Thiele, 2008; Hansen and Tarp, 2001; Jones and Tarp, 2016; Rajan and Subramanian, 2008).[4] These estimators rely on lagged levels and lagged differences of the variables in the model as internal instruments for the contemporaneous equation in differences and levels. Both estimators can be applied to study the additionality of DFI investment, just as they have been used to study aid effectiveness more generally.

Another place to look for evidence of additionality would be in firm-level data. DFIs already 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 can also commission researchers to assemble new firm-level datasets, in order to estimate how the probability of obtaining private investment depends on a project's observable characteristics, and to assess the likelihood that DFI-funded projects would have been able to attract private finance in the absence of the DFI's investment.

3 The data-generating process

The greater emphasis on DFIs in global development cooperation in recent years has been accompanied by new calls for evidence of their additionality. In 2017, for instance, 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 [the UK's DFI], on private sector investment activity" (see also Spratt et al., 2019). In anticipation of donors spending money

[4] Wansbeek (2012) surveys the general popularity of these estimators, while Bazzi and Clemens (2013) discuss their use in growth regressions.

on data-gathering exercises, it is important to ask what these data can and cannot tell us.

To answer this question, we provide a formal representation of DFIs' investing process that we use to generate datasets with known degrees of additionality. We can then systematically investigate the ability of different estimation methods to recover the true extent of additionality. We use the same data-generating process (DGP) for analysing cross-country and firm-level identification strategies, although not all aspects of the DGP will be relevant in each case. We first describe the key features of our DGP, before setting out its formal structure in more detail.

Our DGP has three main steps. In the first step, we create a universe of potential investment projects, each with its own (risk-adjusted) expected return. Investors observe expected returns directly, whereas researchers only observe a noisy proxy of a project's expected return. In the second step, DFIs make their investment decisions. The DFI sector selects projects with expected returns between some lower and upper bound, until its budget is exhausted. In the final step, after DFIs have made their investment decisions, the unfinanced projects turn to private investors for financing.[5] Private investors finance only projects that exceed some minimum expected return.

This DGP is constructed on the blueprint of how DFIs operate, while allowing us to vary the degree of additionality. Setting the lower bound for the expected return that DFIs are willing to consider equal to private investors' minimum required return results in zero additionality ($\beta = 0$), because DFIs target only projects that the private sector would also be willing to invest in. Setting the DFIs' upper bound below or at the private investors' minimum required return, on the other hand, achieves full additionality ($\beta = 1$), as DFIs specifically chase projects with a lower risk-adjusted return than private investors would be willing to invest in on their own.[6] As we discuss later, we can also use this DGP to describe catalytic effects ($\beta > 1$). By re-parameterising the DGP, then, we can generate datasets with different degrees of known additionality.

We now set out the DGP more formally. Throughout our description of the DGP, we specify the default values of parameters in brackets. In our simulations, we vary the values of some parameters to explore the performance of econometric techniques under different circumstances.

We construct datasets with nC (90) countries and T (20) periods. In each period, every country generates a fixed number nI (50) of investment opportunities with an associated expected return.

[5]This assumption about the sequencing of investments is more innocent than it may at first appear, and it is not inconsistent with private investors sometimes beating DFIs to an investment. The DFI sector in our DGP has a finite budget. Hence, even when DFIs and private investors chase the same projects, some of the projects DFIs would be willing to invest in will end up receiving private finance. To assess the additionality of the investments actually made by DFIs, how we think about these privately financed projects in our DGP is immaterial; e.g. one can think that for some of these projects DFIs have lost out to private investors, or that DFIs considered the project but preferred other projects given their finite budget, or even that the project was never put in front of DFIs.

[6]DFIs often stress that they aim to support relatively high-risk projects (see e.g. The Association of European Development Finance Institutions, 2016), and there is some empirical evidence consistent with this claim (Gurara, Presbitero and Sarmiento, 2018; Hainz and Kleimeier, 2012).

Hence, the total quantity of investment opportunities in each period is $nC * nI$ (4500 in our default set-up). For the sake of simplicity and transparency, we assume all projects are of the same size (normalised to 1), and that projects are either wholly financed by a DFI or a private investor. Each investment decision can then be represented as either a zero or a one, and the total quantity of investment is equal to the number of projects that receive financing.[7]

For simplicity, we assume that the observable characteristics (sector, geography, management's track record...) of project p in country i in period t can be fully summarised by a single 'project characteristics' variable, pc_{pit}, which is a noisy proxy for the (risk-adjusted) expected return on investment:

$$er_{pit} = pc_{pit} + e_{pit} \quad \text{with} \quad pc_{pit} \sim \mathcal{N}(\mu_i, \sigma_i^2) \quad \text{and} \quad e_{pit} \sim \mathcal{N}(0, \sigma_e^2) \tag{2}$$

Hence, expected returns are divided into a part that is observable to researchers (pc_{pit}), and a part that is not (e_{pit}). The default value for σ_e is 1.

We suppose that there are three country types: those beyond the investment frontier (e.g. Chad), frontier markets (e.g. Tanzania), and emerging markets (e.g. Vietnam). For each type of country, pc_{pit} is drawn from a different distribution, with mean returns being low, medium, and high, respectively. In our default set-up μ_i is set to $[0, 2, 4]$ for the three types. Standard deviations σ_i are set to 1 by default. We initiate the DGP with half the countries as low-type (low average returns), a third medium-type (medium average returns) and a sixth high-type (high average returns). 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. The transition matrix that governs how country types evolve over time is:

$$\mathbf{P} = \begin{bmatrix} .85 & .10 & .05 \\ .05 & .85 & .10 \\ .05 & .05 & .9 \end{bmatrix} \tag{3}$$

The rows correspond to the type at the start of a period, the columns to the type at the end. So, for example, in each period a low-type country has a 85% chance of staying low-type, a 10% chance of becoming medium-type, and a 5% chance of becoming high-type, and so on for the

[7] We do not concern ourselves with the investments of one DFI possibly crowding out those of another DFI. The question that matters to a donor government considering injecting more capital into its DFI – and to researchers evaluating their impact – should, for the most part, not be substitution between DFIs but the margin between the public and the private sectors. A possible 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. This type of crowding out could be analysed within a similar framework as the one we set out in this paper.

other types. We assume the world is developing, so the probabilities of moving up the hierarchy are higher than the probabilities of regress.

The private sector deems a project worthy of investment if its expected return exceeds the lower bound ps_{min} (default: 2). 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. The parameters we have chosen imply that on average just under 8% of investment opportunities in low-type countries are bankable for private investors, 50% in medium-type countries, and just over 92% in high-type countries.[8]

The lower and upper bounds that delineate the set of eligible DFI investments are denoted dfi_{lo} and dfi_{hi}, respectively. Zero additionality is achieved by setting $dfi_{lo} = 2 = ps_{min}$. 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.

The DFI sector has an exogenously given budget to invest, which changes stochastically over time. In practice, DFI budgets could react endogenously to the number of investment opportunities DFIs are interested in, via new capital contributions from shareholders, the sale of previous investments, and in some cases decisions to borrow on capital markets. An exogenous budget for the DFI sector as a whole is nonetheless a useful starting point, because it is an assumption required for consistency of the supply-push IV estimator, and therefore produces a favourable benchmark for this estimator. Later on, we introduce multiple DFIs with endogenous budgets to permit us to draw out the importance of endogenous DFI budgets for the performance of the supply-push IV estimator.

In the first period, by default we set the budget for the DFI sector equal to 10% of the total number of investment opportunities in that period: $DB_1 = 0.1 * nC * nI$ (equal to $0.1 * 90 * 50 = 450$ for default parameter values). From period 2 onwards, the budget is the nearest integer to:

$$DB_t = DB_{t-1} e^{(drift + \eta_t)} \quad \text{with} \quad \eta_t \sim \mathcal{N}(0, \sigma_{db}^2) \tag{4}$$

Default parameter choices are $drift = 0$ and $\sigma_{db} = 0.15$. In some experiments we will consider a deterministically trending budget by choosing a non-zero value for $drift$ and setting $\sigma_{db} = 0$.

We assume that the overall budget of private investors exceeds the total number of investment opportunities in each year, as is standard in any model with a small open economy with access to global capital markets. As a result, all projects with a sufficient expected return receive financing.[9]

[8] These numbers are taken from a single draw of the simulated data with $nC = 90$, $T = 20$, and $nI = 50000$.

[9] If we would set the private sector budget below the quantity of projects private investors are interested in, a DFI could crowd out private investors from an individual deal but still increase the quantity of investment at a

When the quantity of projects the DFI sector is potentially interested in exceeds the available budget, we assume DFIs pick up projects at random from this eligible set until the budget runs out. This does not imply that DFIs are acting at random, but rather that the outcome looks random from the outside, conditional on projects offering acceptable expected returns. This could occur, for instance, because DFIs also consider the social return on investment, which need not be correlated with the private return within this range. Other possible interpretations are that DFIs do not invest in some projects because the project sponsor prefers a private investor's offer, or because the project was never put in front of a DFI in the first place. Later on, we also consider an alternative DFI project selection process, in which investments are selected on the basis of observable project characteristics.[10]

This DGP is constructed to provide in-many-ways-ideal data for a researcher to determine additionality after the fact. We withhold from the researcher only information about true expected returns, and about DFIs' and private investors' decision rules. Anyone who knew these three things would be able to determine the additionality of each DFI-funded project with perfect certainty, so it is reasonable to impose this handicap. With this information, a researcher would also be able to remove the omitted variable bias in the estimation of β by including the number of projects with an expected return over the private sector threshold as a control variable in equation (1). Everything else is known to the researcher: DFI and private investment in each country-period, and the observable characteristics of every project. In reality, researchers rarely have access even to these data.[11]

In the next two sections we investigate whether conventional estimators can recover the true extent of additionality. To estimate country-level regressions, we aggregate the project-level data generated by our DGP. For each cross-country experiment, the statistics we report are based on 1000 iterations of the DGP. To analyse project-level data, we will only need a single run of the

macroeconomic level. Such capital scarcity may be relevant in some market segments DFIs operate in, but more often a lack of good investment opportunities is the binding constraint (see Rodrik and Subramanian, 2009, for a discussion).

[10]Conversely, in some replications the DFI sector's budget exceeds the number of projects with expected return over two in at least one year. In a zero additionality setting, in such years the DFI sector picks up all investments with a sufficient return, and private investment is zero. Results reported below for experiments with zero additionality are virtually unchanged when we exclude from calculations the small number of replications in which this happens. If we calculate the median ratio of global DFI investment to total private investment across all years in each replication, then the median of this variable across replications in our zero additionality setting with default parameter values is about a quarter, suggesting that private investment generally dominates DFI investment in our replications, as it does in reality. Moreover, results for the zero additionality experiments reported below with a stochastic DFI budget are qualitatively unchanged when we instead randomly redraw the DFI budget in each period as $DB_t = a_t * nC * nI$ with $a_t \sim U(0, 0.1)$, which keeps the size of the DFI sector down without limiting its variation over time too much.

[11]Some DFIs report their investments to the OECD's Creditor Reporting System but it is unclear how complete these data are. For instance, the recipient country and the quantity of investment are sometimes omitted in the project-level data, possibly due to concerns over commercial confidentiality, and some DFIs are missing completely. Some researchers have also privately compiled partial datasets from sources such as DFI annual reports (e.g. Kenny et al., 2018; Massa, Mendez-Parra and te Velde, 2016). Recently, the Institute of New Structural Economics at Peking University has started building a global database of DFIs (Xu, Ren and Wu, 2019).

DGP. We run our simulations in Stata, using `gtools` (Cáceres Bravo, 2019) to improve speed.

4 Evidence from country-level data

Before we turn to the results of our Monte Carlo simulations for country-level data, we explain how the covariate $\widetilde{pc}_{it}$ in equation (1) is constructed. $\widetilde{pc}_{it}$ captures average project characteristics in a country-period after adding measurement error:

$$\widetilde{pc}_{it} = \overline{pc}_{it} + m_{it} \quad \text{with} \quad m_{it} \sim \mathcal{N}(0, \sigma_m^2) \tag{5}$$

If we directly include average project characteristics $\overline{pc}_{it}$ in regressions (corresponding to the case where $\sigma_m^2 = 0$), this control variable predicts the overall level of investment extremely well. For instance, with zero additionality and default parameter values, the mean R^2 across 1000 replications in an OLS regression of I_{it} on only $\overline{pc}_{it}$ exceeds 0.98. The average within R^2 for the corresponding fixed effects regression is 0.97. While it is plausible that researchers running cross-country regressions would include a set of variables that collectively proxy for the underlying determinants of investment, it is implausible that these control variables would predict investment as accurately as $\overline{pc}_{it}$ does in our simulated data. Hence, to add realism, and to illustrate how the bias in the estimated degree of additionality varies with the quality of controls, we often add some measurement error to the project characteristics variable before including it as a control. The larger is σ_m^2, the less bias will be removed by the inclusion of $\widetilde{pc}_{it}$ in our regressions.

4.1 OLS and fixed effects

To provide a methodological baseline, we first examine the performance of OLS and fixed effects (FE) estimation for our DGP with default parameter values. Table 1 shows mean estimates and their standard deviation, calculated across 1000 replications, when there is either no additionality ($\beta = 0$) or full additionality ($\beta = 1$). For each case, we report four results: from a regression that excludes $\widetilde{pc}_{it}$ (columns 1 and 5), and from three regressions that include $\widetilde{pc}_{it}$ with decreasing amounts of measurement error (columns 2-4 and 6-8).

When there is no additionality (columns 1-4), the OLS estimator is consistently upward biased and, at a 5% significance level, a researcher would reject the null hypothesis of zero additionality (H_0: $\beta \leqslant 0$) in all 1000 simulated datasets regardless of the model used (these 100% rejection rates are omitted from the table).[12] The explanation is straightforward. If DFIs pick investment

[12]We estimate equation (1) with standard errors that are robust to heteroskedasticity and clustered by country.

Table 1: OLS and fixed effects results

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
β	0	0	0	0	1	1	1	1
pc	excl.	$\sigma_m^2 = 1$	$\sigma_m^2 = 0.5$	$\sigma_m^2 = 0$	excl.	$\sigma_m^2 = 1$	$\sigma_m^2 = 0.5$	$\sigma_m^2 = 0$
Mean $\widehat{\beta}_{OLS}$	3.50	2.00	0.99	0.21	-2.36	-0.51	0.29	0.75
Std. dev.	0.74	0.25	0.11	0.04	0.77	0.26	0.11	0.04
Mean $\widehat{\beta}_{FE}$	3.03	1.87	0.97	0.21	-1.79	-0.47	0.26	0.73
Std. dev.	0.55	0.22	0.10	0.04	0.55	0.23	0.10	0.04

Note: mean value and standard deviation of OLS and fixed effects estimates of β, based on 1000 replications of our DGP. DFIs randomly select projects from their eligible set. pc indicates when $\widetilde{pc}_{it}$ is excluded ('excl.') or, when it is included, how much measurement error has been added to it.

opportunities "at random" from the same pool as private investors, there will be a spurious positive correlation between DFI and total investment. In country-periods with more plentiful investment opportunities that offer returns over a minimum threshold shared by DFIs and private investors, there will both be more investment by DFIs and more overall investment. As expected, this bias is larger when we do not control for average project characteristics, or when the measure is noisier. The noisier this control variable is, the worse a job it does proxying for the number of projects with sufficient expected return, which is the relevant omitted variable.[13] Fixed effects estimation does not remove the bias since, even conditional on average project characteristics, the number of suitable investment projects in a country varies over time.

In the case of full additionality (columns 5-8), OLS is downward biased, and the null of full additionality (H_0: $\beta \geqslant 1$) is universally rejected in all four specifications. Because DFIs target projects with sufficiently low risk-adjusted expected returns that private investors are not interested in, DFIs and private investors react in opposite ways to shocks to expected returns. An increase in the number of projects with an expected return above the minimum threshold demanded by private investors will lead to more private investment in that country-period. But since DFIs are targeting projects with expected returns in a lower range, there will be less DFI investment. This produces a spurious negative correlation between DFI and private investment, and a downward bias in the estimation of β.

If we assume that, in addition to being fully additional, every unit of DFI investment catalyses an extra two units of private investment, say, so that $\beta = 3$, mean estimates simply shift up by two units, and biases are unchanged. This suggests that, if DFI investment mobilises additional private investment, OLS and FE estimates would underestimate this catalytic effect.

In sum, OLS and FE may be biased upward or downward depending on the (unknown) true degree of additionality. Not much is needed for this bias to manifest itself. Simply having DFIs,

[13]With $\sigma_m = 0.5$ in column 3 the mean R^2 from an OLS regression on just $\widetilde{pc}_{it}$ is still 0.9, and the mean within R^2 in FE estimation 0.85. For $\sigma_m = 1$, in column 2, these numbers are 0.71 and 0.62, respectively.

like private investors, react to expected returns in ways that are not fully observable is sufficient to generate bias.[14]

Next, we investigate whether a supply-push IV estimator and system GMM can provide a solution to this problem. We focus on the zero additionality case with random selection of projects, and examine whether these methods are able to get rid of the upward bias found in OLS and FE estimation. Full additionality results are discussed in appendix A.

4.2 Supply-push IV

The results in Table 1 establish the endogeneity of DFI investment in OLS and FE estimation of equation (1). An obvious solution would be to find an instrument that is both valid and strong, but this is easier said than done. A supply-push instrument, however, that relies on changes in DFIs' total investment budgets affecting some countries (those with strong links to the DFIs in question) more than others, at first glance seems like it could fit the bill.

To fully investigate whether a supply-push instrument is indeed able to provide exogenous variation in DFI investment, we extend our DGP to multiple DFIs (we set the number of DFIs, nD, equal to 3). To avoid having to create an ad hoc method for allocating investments to competing DFIs, we exploit the fact that DFIs sometimes co-invest in the same project. We introduce (time-invariant) DFI preferences over countries, $p_{d,i}$, which are then weighted and used to determine what share of DFI investment in a given country is taken by each DFI. Formally,

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

where d indexes DFIs and i indexes countries. $p_{d,i}$ is between 0 and 1. The DFI with the strongest preference for a country then takes the largest share, $S_{d,i}$, of each DFI investment in that country:

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

The weight ϕ determines to what extent stronger DFI preferences for a country translate into larger shares taken by that DFI of any DFI investment in that country. When $\phi = 0$, all DFI investments are divided up equally between DFIs and there is no meaningful change from having a single DFI. When the weight is large (tending towards infinity) the DFI with the strongest preference takes the entire investment. Setting $\phi > 0$ turns the budget of an individual DFI into

[14] In appendix A.1 we discuss how an alternative selection mechanism, in which DFIs first pick the projects with the worst project characteristics from their eligible set, may change the sign of the bias in the zero additionality case. This illustrates further how the estimated degree of additionality can depend on the specific way in which DFIs attempt to carry out their mandate.

a positive function of the total number of DFI-eligible investment opportunities generated in the countries for which this DFI has a strong preference. Below, we discuss how this endogeneity of DFI budgets creates problems for the supply-push IV estimator.

The supply-push instrument is constructed as follows:

$$dfiIV_{it} = \sum_{d=1}^{nD} s_{i0}^{d} D_{dt} \tag{8}$$

where s_{i0}^{d} is the share of DFI d's total investments that country i receives over an initial period that is excluded from estimation, and D_{dt} is DFI d's budget in period t (the total quantity it invests in that period).[15] The reasoning behind this instrument is that, if a DFI's total investment budget increases for some exogenous reason (say, a political decision in the donor country), countries that are in some sense close to this DFI (i.e. have a high initial share s_{i0}^{d}) will experience an increase in the amount of DFI investment they receive for reasons that should be uncorrelated with their domestic circumstances. The instrument is to be used with a FE IV estimator, which we implement using `xtivreg2` in Stata (Schaffer, 2010). We calculate s_{i0}^{d} as the average share in the first five periods of our samples, setting aside the remaining 15 periods for estimation.

Table 2 shows median estimates and their standard deviations, and rejection rates for zero additionality tests conducted at a 5% significance level (using cluster-robust standard errors). We focus on medians because, in some experiments, outliers pull the mean around. We also report the percentage of replications in which a cluster-robust Kleibergen and Paap (2006) LM test rejects the null of underidentification at a 5% significance level, and the median of a cluster-robust version of the first-stage F-statistic to gauge instrument strength.[16]

As we did before for the OLS and FE estimators, we start in column 1 by applying the supply-push IV estimator to the DGP with default parameter values (which now also include $nD = 3$ and $\phi = 0$). This results in an upward bias. Without further analysis it is, however, unclear whether this is because the instrument is invalid, or because it is weak (the median first-stage F-statistic is low, and underidentification is rejected in only about half of the replications). We are primarily interested in identifying situations where the instrument is invalid; proceeding with

[15] A similar instrument is often used to estimate the labour market effects of immigration (e.g. Card, 2001). Compared to the canonical shift-share instrument proposed by Bartik (1991) and analysed recently by, among others, Borusyak, Hull and Jaravel (2018) and Goldsmith-Pinkham, Sorkin and Swift (2019), the shares used in the construction of our instrument are calculated in a different way. Also, the fact that the shifter in our case is a budget that is distributed across recipient countries has implications for the performance of the leave-one-out version of the instrument, which we will come to later.

[16] To test the null of weak instruments, it is standard to compare this F-statistic to the critical values in Stock and Yogo (2005). This could be misleading since these critical values rely on iid errors (see e.g. Bun and de Haan, 2010). Olea and Pflueger (2013) propose a robust test for weak instruments but in our case, with a single endogenous variable and a single excluded instrument, their test statistic equals the robust first-stage F-statistic (though it would have to be compared with larger critical values than those computed by Stock and Yogo).

Table 2: Supply-push IV results with zero additionality ($\beta = 0$)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Δ types	default	fewer	1period	fixed	1period	1period	1period
ϕ	0	0	0	0	1	2	2
pc	excl.	excl.	excl.	excl.	excl.	excl.	$\sigma_m^2 = 0.5$
σ_{db}	0.15	0.15	0.15	0.15	0.15	0.15	0.05
Med. $\widehat{\beta}_{IV}$	4.56	0.53	0.00	0.00	0.12	0.26	0.39
Std. dev.	84.72	19.65	0.97	0.09	0.63	0.54	0.53
% reject $\beta \leqslant 0$	68.6	43.3	8.2	4.7	15.9	27.4	29.7
Med. F	4.77	23.6	37.1	166	43.7	49.4	20.5
% reject underid.	53.4	72.9	97	100	98.8	99.7	97.4

Note: median value and standard deviation of IV estimates of β, based on 1000 replications of our DGP. % reject $\beta \leqslant 0$ is the percentage of replications in which the null of zero additionality is rejected at a 5% significance level. The final two rows show the median cluster-robust first-stage F-statistic, and the percentage of replications that reject underidentification at a 5% significance level. DFIs randomly select projects from their eligible set. $\Delta types$ states how often types change over time: as given by transition matrix (3) ('default'); reduced probabilities of transitions as in transition matrix (9) ('fewer'); transitions occur for one period only ('1period'); types are fixed ('fixed'). pc indicates when $\widehat{pc}_{it}$ is excluded ('excl.') or, when it is included, how much measurement error has been added to it.

a weak instrument in practice is likely to lead to bias and imprecise estimates, but at least it is possible to verify instrument strength empirically. In order to separate weak instrument bias from bias caused by instrument endogeneity, we first tweak the DGP so that the instrument becomes strong enough for there to be no weak instrument bias.

In columns 2-4, we increase instrument strength by gradually making country types more persistent. In column 2, we reduce the probabilities of changing types by implementing the following transition matrix instead of the one given in (3):

$$\mathbf{P} = \begin{bmatrix} .97 & .02 & .01 \\ .01 & .97 & .02 \\ .01 & .01 & .98 \end{bmatrix} \tag{9}$$

In column 3 the probabilities in transition matrix (3) apply, but transitions only occur for a single period; at the start of the next period a country always reverts to its core type. So, for instance, a country that is low-type at its core has, in each period, a 0.85 probability of being low-type, a 0.1 probability of being medium-type, and a 0.05 probability of being high-type. In column 4, types are fixed. As types become more persistent, the instrument becomes stronger and the bias disappears. Restricting transitions to a single period in column 3 is sufficient to remove all bias, so we continue with this transition mechanism for the remainder of Table 2.[17]

In what follows, we discuss two reasons as to why the supply-push instrument may not yield

[17] We do not necessarily see this transition mechanism as being realistic. We use it because it is a simple way to obtain a benchmark with a strong instrument and zero bias, from which we can explore the circumstances in which the supply-push instrument becomes invalid.

reliable estimates of the degree of additionality, even when the instrument is strong. The first has to do with endogenous reactions of DFI budgets to the number of investment opportunities, the second with trends in the global DFI budget combined with changes in country types.

We first consider bias induced by endogenous reactions of DFI budgets. It is reasonable to suppose that DFIs have persistent preferences for some countries over others, caused by historical ties or explicit strategy decisions (or by the regional focus of regional development banks), and that the quantity of investment DFIs make responds to whatever investment opportunities arise in these countries. Under zero additionality, if the countries they favour experience an improvement in their investment climate, DFIs are likely to react by increasing their total investment. Unlike aid agencies that largely disburse grants, DFIs do not really have a predetermined budget to spend each year. Rather, they have access to financial resources which, to an extent, they can draw down in response to demand. Hence, the total quantity a DFI invests in any given period is not necessarily exogenous to changes in circumstances in countries a DFI has a preference for. This mechanism contaminates the supply-push instrument, so that it may yield an inconsistent estimator of DFI additionality.

To investigate this, we continue with the set-up from column 3 where deviations from core type only occur for a single period, but increase ϕ first to one (column 5), then to two (column 6). When $\phi = 0$, DFI-specific preferences do not matter, and any DFI investment, regardless of where it takes place, is split equally among the three DFIs. As a result, each DFI's total investment budget is a third of the sector's overall budget DB_t and hence determined by a completely exogenous process. In contrast, when $\phi > 0$, a DFI takes up a larger share of DFI investments in countries that it has a stronger preference for. This makes individual DFI budgets endogenous. If a country experiences an unobserved positive shock to the number of high-return projects, it will attract more DFI investment. A DFI with a strong preference for this country will take a large share of this extra DFI investment, and see its overall investment budget rise as a result. Hence, when $\phi > 0$, an individual DFI's budget becomes a positive function of unobserved shocks to expected returns in the countries it has a strong preference for. This invalidates the instrument: an unobserved positive shock to expected returns increases the value of the instrument in that country-period, since D_{dt} increases most for those DFIs that have the strongest link (largest s_{i0}^{d}) with the country that has experienced the positive shock. This creates a positive correlation between u_{it} and $dfiIV_{it}$.[18] In column 5, with $\phi = 1$, the IV estimator is upward biased, and this bias grows in column 6 when we increase ϕ to two. This bias is purely due to the instrument

[18]Related to this point, Borusyak, Hull and Jaravel (2018) discuss the general importance of the shifter being uncorrelated with a weighted average of unobserved shocks for consistency of the canonical shift-share IV estimator.

becoming invalid, as it is stronger than in column 3, were no bias is found.[19]

In our DGP, we assume that the global DFI budget DB_t is exogenous, to give us a benchmark in which the supply-push instrument is valid. In reality, the DFI sector's total investment may also respond to the number of projects DFIs are interested in. We discuss this in appendix A.2.1, where we also argue that endogenous responses of DFI budgets likely lead to a trade-off between instrument strength and validity.

An example of the endogenous response of DFI budgets can be seen in the counter-cyclical role sometimes played by DFIs, especially in times of crisis (for discussions, see e.g. FMO Development Impact Team, 2014; Griffith-Jones, 2016; Independent Evaluation Group, 2008; Spratt and Collins, 2012; te Velde, 2011). Massa, Mendez-Parra and te Velde (2016) note, for instance, that, over the period 2013-15, the European Investment Bank provided €60 billion in additional lending to aid the recovery of Europe during the eurozone crisis. In the full additionality version of our DGP, this could be interpreted as a DFI that wants its investments to be additional expanding its overall budget to take advantage of an increased number of investment opportunities that the private sector would not be interested in (e.g. because the crisis has lowered private investors' assessment of risk-adjusted expected returns).

Controlling for $\widetilde{pc}_{it}$ again reduces bias. It is, however, easy enough to generate situations where, even with $\widetilde{pc}_{it}$ included, substantial bias remains. We give an example in column 7, for a small amount of measurement error added to $\widetilde{pc}_{it}$ ($\sigma_m^2 = 0.5$). We keep $\phi = 2$ and shrink the time-series variation in the DFI sector's overall budget by decreasing σ_{db} from 0.15 to 0.05. This change weakens the instrument (the median first-stage F-statistic falls from about 50 to 20), which exacerbates bias. The bias is close to 0.4, which almost doubles when we add more measurement error to $\widetilde{pc}_{it}$ ($\sigma_m^2 = 1$, not reported). In appendix A.2.2 we further show that a leave-one-out version of the instrument, commonly employed in shift-share instruments to deal with feedback from an individual unit to the aggregate shifter, does not remove the bias.

These results suggest it would be fruitful to find natural experiments that allow researchers to identify exogenous variation in DFI's investment budgets. Replacing the total amounts invested by DFIs in the construction of the supply-push instrument by an exogenous predictor for these amounts that is not influenced by recipient country circumstances would get rid of the problems caused by endogenous DFI budgets. In the absence of compelling exogenous variation in DFI's total investment budgets, showing that there is a large degree of fragmentation on both the

[19]Switching on the DFI-specific preferences for recipient countries by setting $\phi > 0$ provides an additional reason for persistence in DFIs' allocation of investments. However, the resulting increase in instrument strength is muted compared to the changes in instrument strength from making types more persistent in the first four columns.

recipient and DFI side could provide useful support for the instrument's validity.[20]

Table 3: Supply-push IV results with zero additionality ($\beta = 0$): downward trend in the DFI budget

	(1)	(2)	(3)	(4)	(5)
drift	-0.1	-0.1	-0.1	-0.1	-0.1
Δ types	default	fewer	fixed	default	default
i.trend	no	no	no	yes	no
pc	excl.	excl.	excl.	excl.	$\sigma_m^2 = 0.5$
Med. $\widehat{\beta}_{IV}$	8.43	2.91	0.00	1.07	2.15
Std. dev.	1.75	0.79	0.10	2.87	1.12
% reject $\beta \leqslant 0$	99.2	99.3	5.4	17.6	77.9
Med. F	29.4	111	167	12.9	15.3
% reject underid.	99.6	100	100	99.7	96.6
Med. FS	0.71	0.97	0.95	1.33	0.41
Med. RF	5.99	2.83	0.00	1.39	0.89

Note: see Table 2. This table also shows median first stage ('FS') and reduced form ('RF') estimates. For all results in this table $\sigma_{db} = 0$. i.trend indicates in which columns time dummies have been replaced by country-specific trends.

A second reason for an invalid supply-push instrument, explored in Table 3, is the combination of trends in the global DFI budget and countries changing types.[21] We explain this bias for a downward trending budget, which shows the mechanics at play most clearly. The case with an upward trending budget is similar, but more complicated, so we discuss it in appendix A.2.3. In the discussion that follows, it is useful to keep in mind that our IV estimator can be written as the ratio of the reduced form coefficient and the first stage coefficient.

In Table 3 we return to the default version of our DGP that we also used in the first column of Table 2 (with $nD = 3$ and $\phi = 0$), with one change: we remove the stochastic element in the DFI budget by setting $\sigma_{db} = 0$, and replace it with a deterministic downward trend ($drift = -0.1$). In column 1 we consider default type transitions, while column 2 examines what happens with reduced probabilities of type changes (as given by transition matrix (9)). When the global DFI budget trends downwards, the IV estimator is upward biased in both cases. In both experiments, the first stage coefficient is positive in every single replication: because of the downward trend in the global DFI budget, DFI investment declines more sharply over time for countries that start out as high-type than for countries that are low-type in the first period, and the same goes

[20]See Temple and Van de Sijpe (2017) for a detailed discussion. Roughly speaking, if DFIs spread their investment budgets relatively evenly over a large number of countries, then it is less likely that many DFI budgets are unduly influenced by the changing circumstances in just a handful of countries. Similarly, if instrument values for each country depend on a large number of DFIs and the overall budgets of these DFIs do not all react in the same way to shocks to expected returns, then the correlation between the instrument and the error term is weakened. The role played by fragmentation is evident from the results we reported earlier when increasing ϕ, as larger values for ϕ correspond to less fragmentation.

[21]This point is related to the analysis in Christian and Barrett (2019), who mostly focus on a situation where the true coefficient of an interacted instrument in the first stage is zero, and stochastic trends in the time series component of the instrument, the endogenous variable, and the outcome generate spurious correlations that bias the IV estimator.

for $dfiIV_{it}$. As a result, the instrument tracks actual DFI investment well, resulting in positive first stage coefficients and a strong instrument. The reduced form coefficient as well is positive in every replication: because of changes in types, a country that starts out as high-type sees total investment fall when its type shifts downwards, while a country that is initially low-type is more likely to experience increases in total investment. $dfiIV_{it}$ falls more rapidly for initial high-type countries than for countries that start as low-type, producing positive reduced form coefficients. As both the reduced form and first stage coefficients are positive in each replication, the IV estimator is upward biased. This bias is less pronounced in column 2 because the reduced probability of type changes imply that changes in total investment over time are less pronounced, which pushes the reduced form coefficient closer to zero. When types are fixed (column 3) there are no differential changes in total investment by initial type, resulting in a median value of zero for the reduced form coefficient and for the IV estimator.

Column 4 shows that, for the default transition mechanism, replacing time dummies by country-specific trends is insufficient to fully remove this bias. Likewise, controlling for $\widetilde{pc}_{it}$ (with $\sigma_m^2 = 0.5$) in column 5 reduces the bias but does not eliminate it.

The supply-push instrument offers an intuitive way to instrument for the amount of DFI investment received, but our results suggest that this approach is not without risk.[22] The instrument could yield misleading results if the total quantities invested by DFIs respond to changes in the number of investment opportunities of interest to DFIs. Documenting a large degree of fragmentation in DFI investment, and especially finding an exogenous predictor for DFI budgets, could help to alleviate this concern, but, even then, the combination of trends in the global DFI budget and changing investment environments would result in unreliable estimates of DFI additionality.

4.3 System GMM

In the absence of valid external instruments, difference and system GMM estimation instead rely on internal instruments for identification. Difference GMM (Arellano and Bond, 1991; Holtz-Eakin, Newey and Rosen, 1988) starts by differencing equation (1) to remove w_i:

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

followed by using suitably lagged levels of the variables as instruments within Hansen's (1982) GMM framework. System GMM (Arellano and Bover, 1995; Blundell and Bond, 1998) further adds the equation in levels (equation (1)), instrumenting it with lagged differences of variables

[22] The results for full additionality discussed in appendix A.2.4 mirror the zero additionality results.

(see Bond, 2002; Bun and Sarafidis, 2015; Roodman, 2009*a*, for excellent introductions). We implement these estimators using the `xtabond2` command in Stata developed by Roodman (2009*a*).

We calculate one-step GMM estimates with cluster-robust standard errors. We treat dfi_{it} and, when included, $\widetilde{pc}_{it}$ as endogenous. To avoid overfitting (Roodman, 2009*b*), we use only a single lagged level of each variable as an instrument for the differenced equations. This yields the following population moment conditions:[23]

$$
\begin{aligned}
&\mathrm{E}\left[dfi_{i,t-2}\Delta u_{it}\right] = 0 && \mathrm{E}\left[\widetilde{pc}_{i,t-2}\Delta u_{it}\right] = 0 \\
&\mathrm{E}\left[\Delta dfi_{i,t-1}\left(w_i + u_{it}\right)\right] = 0 && \mathrm{E}\left[\Delta\widetilde{pc}_{i,t-1}\left(w_i + u_{it}\right)\right] = 0
\end{aligned}
\tag{11}
$$

To further counter overfitting, we also collapse the instrument matrix (Roodman, 2009*a*). Time dummies are used as instruments in the levels equation only; their use as instruments in the differenced equation is redundant. When the moment conditions in (11) hold, GMM is consistent but not unbiased. Non-negligible bias can result from violated moment conditions or from weak instruments, or from a combination of both.

Table 4 reports median system GMM estimates of β, and their standard deviations. We also include rejection rates for a (one-sided) t-test of the null of zero additionality, conducted at a 5% significance level. Hansen % pass is the percentage of replications that do not reject Hansen's overidentifying restrictions test at a 10% significance level. The ability of this test to pick up moment violations is hampered, however, by its low power (Bowsher, 2002). It also starts from the assumption that there are enough valid moment conditions to identify the model's coefficients; if all moments are violated in similar ways, this test is unlikely to reject.[24]

Finally, we carry out a test for underidentification. We report the cluster-robust version of the Sanderson and Windmeijer (2016) conditional first-stage F-statistic proposed in Windmeijer (2018). This test assesses whether the instruments are strong enough to identify the parameter of interest, β, specifically. This is the most relevant available test statistic for our purposes, as Sanderson and Windmeijer (2016) show that, when there are multiple endogenous variables and some are instrumented weakly, the coefficients of the variables that are instrumented strongly are still estimated consistently. We include the median value of the test statistic, as well as the percentage of replications that reject the null of underidentification at a 5% significance level.

[23]Results using lagged levels in $t-2$ through to $t-5$ as instruments for the differenced equations are qualitatively similar, with larger biases.

[24]The difference-in-Hansen tests we conducted to shed light on the validity of specific subsets of moments are not very informative, so we do not report them. The same goes for Arellano and Bond's (1991) $m2$ test, whose results do not vary much across experiments and tend to indicate no serial correlation in u_{it} in the vast majority of replications.

Table 4: System GMM results with zero additionality ($\beta = 0$)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
LDV	no	no	no	no	yes	yes	yes	yes
pc	excl.	$\sigma_m^2 = 1$	$\sigma_m^2 = 0.5$	$\sigma_m^2 = 0$	excl.	$\sigma_m^2 = 1$	$\sigma_m^2 = 0.5$	$\sigma_m^2 = 0$
Med. $\widehat{\beta}_{sysGMM}$	3.59	0.58	0.15	0.02	0.50	0.42	0.20	0.02
Std. dev.	0.97	1.44	0.96	0.35	1.55	1.11	0.80	0.30
% reject $\beta \leqslant 0$	100	25.6	13.3	5.3	24.2	20.5	12.3	6.4
Hansen % pass	33.3	86.8	91.2	91.7	80.8	90	91.5	91.7
Med. cond. F	32.7	8.92	13.8	17.9	11.3	7.92	10.4	17.2
% reject underid.	100	54.8	69.7	76	64.8	42.8	53.4	72.1

Note: median value and standard deviation of system GMM estimates of β, based on 1000 replications of our DGP. % reject $\beta \leqslant 0$ is the percentage of replications in which the null of zero additionality is rejected at a 5% significance level. Hansen % pass is the percentage of replications that do not reject Hansen's overidentifying restrictions test at a 10% significance level. The final two rows show the median cluster-robust conditional F-statistic, and the percentage of replications that reject underidentification at a 5% significance level. DFIs randomly select projects from their eligible set. LDV indicates whether a lagged dependent variable is included. *pc* indicates when $\widetilde{pc}_{it}$ is excluded ('excl.') or, when it is included, how much measurement error has been added to it.

Table 4 shows results for the same default version of our DGP that was used in the first four columns of Table 1 (reverting to a single DFI: $nD = 1$). In the regression without $\widetilde{pc}_{it}$ (column 1), underidentification is rejected in every replication, but the Hansen test also often rejects, and system GMM is unable to remove the upward bias in the estimation of β. The reason for this is that, in our DGP, the moment conditions in (11) are not satisfied. The main culprit for this are changes in country types.

First consider the moment conditions associated with the differenced equation. A country that is low-type in $t-2$ will receive little DFI investment in this period, because it generates few projects with a sufficient expected return. For this country, the only way is up: it either remains low-type, or it moves up a type (or two), leading to increases in the number of investable projects and the amount of DFI investment. The converse applies to high-type countries. In the real world, a country with few appealing investment projects will not be much affected if its investment climate remains unchanged or even further deteriorates, whereas an improvement in its investment climate will increase the number of projects that are attractive to investors seeking high returns. The consequence is that $\text{Corr}\left(dfi_{i,t-2}, \Delta dfi_{it}\right) < 0$ but also that $\text{Corr}\left(dfi_{i,t-2}, \Delta u_{it}\right) < 0$, where Δu_{it} contains the change in the number of projects with a sufficient expected return as an omitted variable in the differenced equation.[25]

A similar story applies to the levels equation. If a low-type country moves up types in $t-1$, its DFI investment increases, and, since types are persistent, it is also likely to end up with a larger

[25] From the probabilities in transition matrix (3) it can easily be verified that, for a country that is low-type in $t-2$, the likelihood of moving up a type between $t-1$ and t (so that $\Delta u_{it} > 0$) exceeds the likelihood of moving down a type. Likewise, for a high-type country in $t-2$, a downward shift in type between $t-1$ and t is more likely than an upward shift. The correlations discussed in the text can be calculated from the data generated by our DGP, since we can measure $w_i + u_{it}$ as the number of projects with sufficient expected returns, and Δu_{it} as the change in this variable.

number of projects with high expected returns in period t, implying that $\text{Corr}\left(\Delta dfi_{i,t-1}, dfi_{it}\right) > 0$ but also that $\text{Corr}\left(\Delta dfi_{i,t-1}, w_i + u_{it}\right) > 0$. Trends in the DFI sector's budget can also give rise to violations of the moment conditions in the levels equation, even when types are time-invariant. For instance, if the global DFI budget trends upwards, high-type countries benefit most from this, generating a positive correlation between $\Delta dfi_{i,t-1}$ and w_i.[26]

As was the case for the other estimators, including $\widetilde{pc}_{it}$, especially without measurement error, reduces bias (columns 2-4 in Table 4). $\widetilde{pc}_{it}$ partially controls for the number of projects with sufficient expected returns, weakening the correlations between instruments and error terms. Table 4 makes clear, however, that there is nothing inherent in system GMM that removes the bias in the estimation of β. The good performance of the estimator in column 4 depends on the availability of a control variable that almost perfectly predicts where investments will take place. Without such a control, system GMM clearly returns a bias.

In the final four columns of Table 4 we show that this conclusion holds when we add $I_{i,t-1}$ as a covariate. The inclusion of a lagged dependent variable is typical in system GMM estimation; one reason for this is to remove serial correlation in u_{it}, which would otherwise invalidate the moment conditions. As is common, we treat $I_{i,t-1}$ as predetermined, exploiting moment conditions $\text{E}\left[I_{i,t-2}\Delta u_{it}\right] = 0$ and $\text{E}\left[\Delta I_{i,t-1}\left(w_i + u_{it}\right)\right] = 0$. The main change from including $I_{i,t-1}$ is lower bias in the model without $\widetilde{pc}_{it}$.

Our DGP creates a forgiving test-bed for difference and system GMM estimators, yet relatively persistent shifts over time in a country's average expected returns are enough for moment conditions to be violated and for these estimators to yield unreliable results.[27] Adding other realistic features, like serially correlated shocks to a country's expected returns or multiple variables to measure project characteristics, would likely undermine their performance even further. In any realistic setting, where countries' investment climates change systematically over time, the GMM estimators considered in this paper do not provide reliable estimates of the degree of additionality.

5 Evidence from firm-level data

Another place to look for evidence of additionality could be observational firm-level data. To estimate the degree of additionality, one could imagine an experiment that awards DFI financing

[26]Appendix A.3.1 describes a number of additional experiments to illustrate how these moment violations come about.

[27]The full additionality results in appendix A.3.2 are the mirror image of the zero additionality results, showing downward biases.

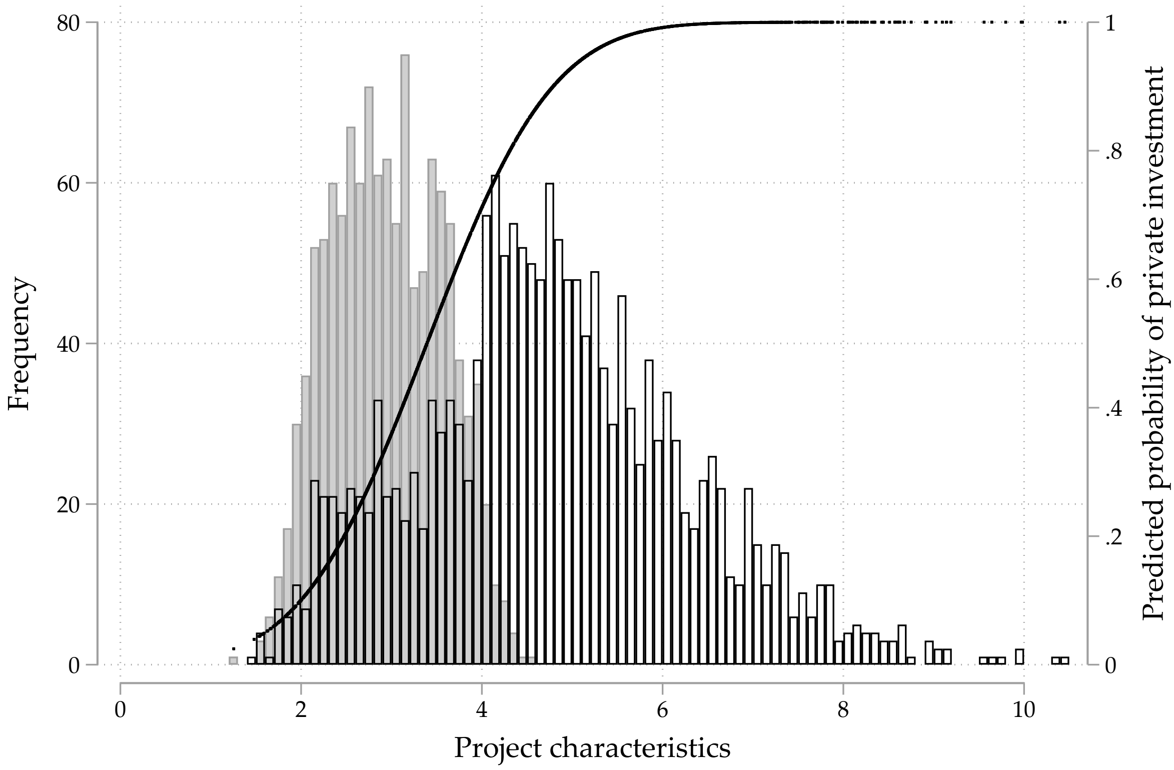
to a randomly chosen subset of prospective investees and tracks the funding outcomes of rejected projects, but this is unworkable in practice. The average deal size reported by The Association of European Development Finance Institutions (2016) was just under $9 million, so it is highly unlikely that anyone would run a trial with amounts of that size invested at random. 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. We therefore focus on what can be learned from observational firm-level data. The firm-level data will not speak to questions about general equilibrium effects of DFI investments but, by allowing researchers to compare the observable characteristics of DFI-funded and privately funded projects, they can perhaps establish the likelihood of DFI-funded projects being additional.

Before investigating the usefulness of such an approach, we first note that, while there is a sound reason for looking at the level of investment as an outcome in aggregate data, this logic does not carry over to firm-level data. Firms financed by DFIs may have greater or smaller levels of investment than privately financed firms, irrespective of the degree of additionality. If a regression was run and revealed that DFI investees on average invest more than comparable firms that are financed privately, we would not learn anything about additionality, since this approach does not tell us whether these DFIs displaced private investors. It might be the case, for instance, 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. To identify investment additionality a different approach is therefore required.

For the purposes of examining the degree of additionality, the outcome of interest in firm-level data is whether the funder is a DFI or a private investor. To see what firm-level data might reveal about additionality in a best-case scenario, let us suppose that the 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 expected return for every single project (this is our project characteristics variable). Would this allow the researcher to identify the true extent of DFI additionality?

A natural way to assess additionality using these data would be to estimate a discrete choice model where the funder's identity is a function of project characteristics, in the hope that this reveals 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 funded by the private sector. If these probabilities are high, we might infer that the degree of additionality is low, and vice versa. An example is presented in Figure 1, which shows the frequency of DFI and private investment for different values of project characteristics in a simulated dataset.[28] It also plots the predicted probability of receiving private funding as a function of project characteristics, based on a simple probit regression that is run on the sample of projects that receive either DFI or private finance.

Figure 1: Inferring additionality from firm-level data



Note: gray bars for DFI-funded projects, transparent bars with black lines for privately funded projects. The black dots show the predicted probability that a project is privately funded rather than DFI-funded, as read on the right-hand y-axis. Parameters to generate the data are: $nC = 12$ (4 of each type), $T = 1, nI = 500, DB = 0.2 * nC * nI, \mu_c = [0, 2, 4], \sigma_c = [2, 2, 2], \sigma_e = 0.25, dfi_{lo} = ps_{min} = 2, dfi_{hi} = 4$, random selection mechanism. This graph uses the `plotplainblind` scheme provided by Bischof (2017).

This plot looks exactly as DFIs would hope. The bulk of DFI investments have project characteristics in the low range, where the predicted probability of private investment is quite low. The predicted probability that the private sector would have undertaken an investment climbs to 0.5 for projects that have characteristics towards the top end of DFIs' range. A researcher could conclude that most DFI investments are probably additional, although there is a good chance (> 0.5) that roughly a third of investments, with project characteristics towards the higher end of

[28]The data are generated assuming that 12 countries, four from each type, are observed for a single period, each with 500 investment opportunities. A full parametrisation is given in the note to the figure.

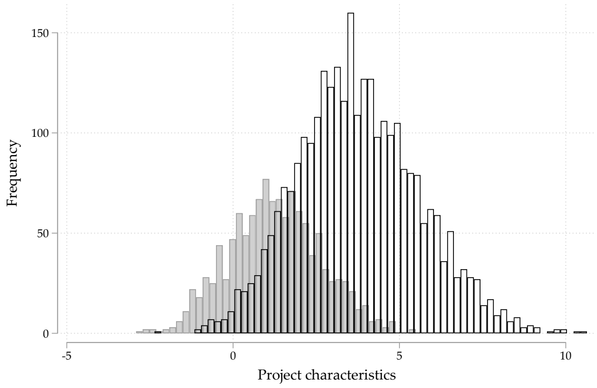
the DFI range, crowded out private investors.

This conclusion would be incorrect, however. The encouraging results from Figure 1 were generated by a DGP with zero additionality. Because the DFI sector's budget is relatively large, however, DFIs pick up most of the projects within their investment band, making it seem as if private investors are not interested in these projects. The fundamental problem with the analysis in Figure 1 is that it conflates what private investors did in the presence of DFIs with what they would have done in their absence. To make 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.

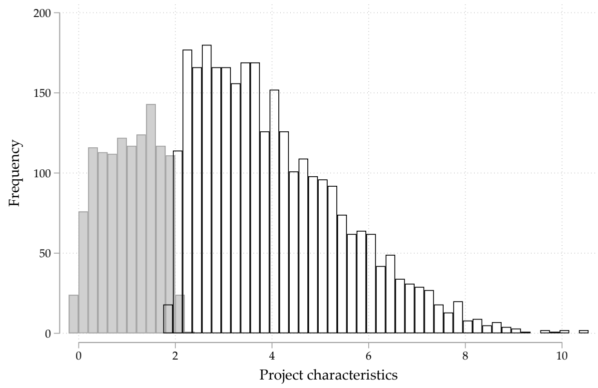
Suppose private investors had been presented with all of the projects in our dataset and then rejected some that were subsequently funded by a DFI. In that model, the fact that a project is DFI-funded is indeed revealing about its additionality: it is additional by construction. If, on the other hand, DFIs select their investments first, then DFI-funded projects are removed from private investors' choice set, severely limiting the ability of the data to reveal whether private investors would have invested in these projects. From looking at Figure 1, there is no way to tell whether private investors have little interest in projects with project characteristics in the DFIs' investment range, or whether a well-funded DFI sector has pipped private investors to the post for these projects.

Figure 2 gives four additional examples to further illustrate how this approach can yield misleading results. Both plots in the top row are from a DGP with full additionality. Nonetheless, in Figure 2a it looks as if there is a lot of overlap in the type of projects DFIs and private investors are interested in, from which a researcher might erroneously conclude that additionality is low. This is because in this DGP we have assumed a large variance for the unobserved component of expected returns ($\sigma_e = 1.5$), so that project characteristics are a less good guide to expected returns. When the variance is lower ($\sigma_e = 0.1$), as in Figure 2b, the full additionality of DFI investments reveals itself clearly. In the bottom row, the DGP has zero additionality. In Figure 2c DFIs' budget is large enough ($DB = 0.3 * nC * nI$) for them to pick up all the projects they are interested in, which makes it look practically indistinguishable from Figure 2b. As in Figure 2b, a researcher would conclude that DFI investment is fully additional, but would in this case be wrong. In contrast, when the DFI budget is curtailed ($DB = 0.1 * nC * nI$ in Figure 2d), the researcher is likely to correctly conclude that DFI investment is not additional. That two completely different additionality scenarios can give rise to similar patterns in the data (Figures 2b and 2c, as well as Figures 2a and 2d) clearly shows the problems associated with inferring

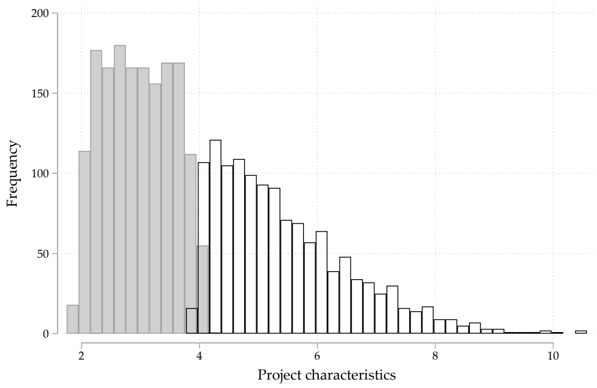
Figure 2: Inferring additionality from firm-level data: additional examples



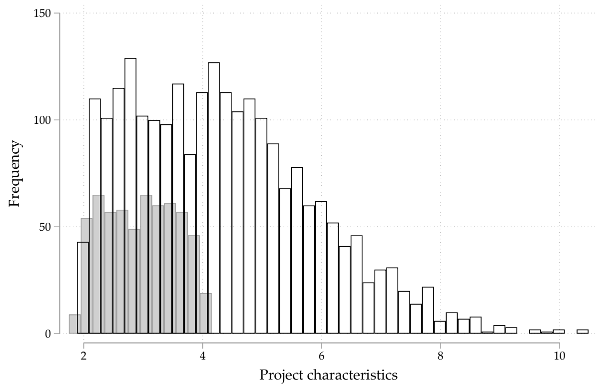
(a) Full additionality, large errors



(b) Full additionality, small errors



(c) Zero additionality, DFI funds > opportunities



(d) Zero additionality, DFI funds < opportunities

Note: see Figure 1. All figures share the following parameter values: $nC = 12$ (4 of each type), $T = 1, nI = 500, \mu_c = [0,2,4], \sigma_c = [2,2,2], ps_{min} = 2$, random selection mechanism. In the top row there is full additionality: $dfi_{lo} = 0, dfi_{hi} = 2$. In Figure 2a the variance in the non-observable part of expected returns is higher ($\sigma_e = 1.5$) than in Figure 2b ($\sigma_e = 0.1$); $DB = 0.2 * nC * nI$ in both figures. In the bottom row there is zero additionality: $dfi_{lo} = 2, dfi_{hi} = 4$. In Figure 2c the budget is larger ($DB = 0.3 * nC * nI$) than in Figure 2d ($DB = 0.1 * nC * nI$); $\sigma_e = 0.1$ in both figures.

additionality from firm-level data.

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. DFI investments may look different from private investments, but that could be because DFIs are taking all available deals of a certain type, even though private investors would have done them. DFI investments may resemble private investments, yet still be additional because of differences between what insiders base their decisions on and observable project characteristics such as country, sector and investment type. Even with excellent firm-level data, rigorous evidence of additionality may continue to elude researchers.

6 Qualitative evidence

In practice, claims of additionality are often supported by qualitative evidence (see e.g. Analysis for Economic Decisions, 2016; Independent Evaluation Group, 2008; Spratt and Collins, 2012). Given the challenges in providing credible quantitative evidence, let us consider this alternative line of evidence more seriously.

The arguments that additionality is present at the time that the decision to invest is made are the primary type of qualitative evidence, although DFIs also conduct retrospective surveys of their investees and private investors that ask questions about additionality. DFIs' evaluations of additionality tend to be broad, encompassing both investment (quantity) additionality and value (quality) additionality, where the latter refers to a DFI's potential for changing the nature of a project so that it becomes more beneficial or has a higher chance of success. 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" (International Finance Corporation, 2013, p. 4). A technical report for the European Union defines additionality as "the net impact of an intervention after taking into account what would have happened in the absence of the intervention (reference case)" (The European Union blending and external cooperation (EUBEC) platform, 2013, p. 34). 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 (ibid, p. 35). When evidence for the additionality of EU blending is evaluated, it consists of assessing the quality of those arguments (Analysis for Economic Decisions, 2016).

One difference between how DFIs approach additionality in practice and the strict interpretation of investment 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. Finding evidence of financial additionality is, however, not sufficient to establish investment additionality: a DFI may crowd out private investors by offering finance on more favorable terms than private investors would. For example, DFIs may observe that they offer longer-term loans than 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 in this case 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 – could still be more appealing to project sponsors than what private investors can offer, these DFIs may still be crowding out private finance.

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, as their remuneration may depend at least in part on deal volume and returns. 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.

Surveys of project sponsors and competing private investors may be informative, but cannot be entirely relied upon. Although some project sponsors may come to learn that only a DFI would support them, 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; for instance, if given examples of successful DFI investments they may be prone to claiming they would have made what looks like the right decision (of investing) in retrospect.

Process tracing could provide a useful way to organise the available circumstantial evidence of an investment's additionality, or lack thereof. Informally, in process tracing 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.[29] One potential problem is that the scope of process tracing to narrow the bounds on the probability that an outcome can be attributed to an intervention is often limited, even in favourable circumstances (Dawid, Humphreys and Musio, 2019). This may be an especial concern in assessing DFI additionality, as the investment process has relatively few moving parts so that there simply might not be enough observable factors to reach much of a conclusion one way or another.

7 Discussion

Development Finance Institutions are playing an increasingly important role in international development, and with that attention has come increasing demands for evidence that their investments are additional. Unfortunately, our analysis suggests that it may be almost impossible to reliably estimate DFIs' investment additionality from datasets of the kind that researchers are likely to have access to. By articulating the problem as a formal data-generating process that captures the key aspects of how DFIs and private investors interact, we have been able to test a range of estimators on datasets with known levels of additionality. The biases in these estimators generally follow from core features of this process, which suggests that rigorous quantitative evidence of additionality is likely to remain elusive even with more sophisticated techniques and better data. Meanwhile, the type of qualitative evidence commonly used to support claims of additionality is no less immune to bias.

Given the difficulties in reliably establishing a DFI's investment additionality, the most productive course of action may be for DFIs to embrace this fundamental uncertainty as part of their decision making. One concrete proposal is for DFIs to abandon binary assessments of additionality and instead to evaluate additionality probabilistically, as some DFIs already do. DFIs could employ process tracing or related approaches to prospectively assess the probability of additionality. This involves pre-specifying conditions in which they think additionality is more or less likely, and coming up with a weight to these different factors. The qualitative evidence will undoubtedly be flawed, but may nevertheless be sufficient to plausibly shift beliefs. For instance, if a DFI invests in a market that previously has only seen private investment, and it invests in a project with a very different profile and expected return from the existing private investments in this market, then additionality is clearly more likely. In essence, the claim of additionality then

[29] See Collier (2011) for an introduction, Humphreys and Jacobs (2015) and Befani and Stedman-Bryce (2017) for more quantitative Bayesian applications, and Spratt et al. (2019) for a discussion of this method in the context of DFI mobilisation of private investment.

comes down to arguing that we are in the situation described in Figure 2b rather than that of Figure 2c. Boards of directors and shareholders may not be ready to hear: "we think there is 50 per cent chance this project is additional." Yet even if additionality is a binary, one may form a subjective estimate of its probability. At the time of writing, we are aware of four DFIs that rate ex-ante expected additionality on a scale, CDC being a recent convert.

One important advantage of evaluating investment additionality probabilistically is that it can then easily be traded off against other objectives, such as value additionality. For the sake of illustration, suppose a DFI confers value 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. 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 $\Delta + \pi_a \Omega$. A lower probability of additionality could then be traded off against larger anticipated benefits conditional on investment additionality and/or by greater value additionality. A pre-specified process of gathering and weighting qualitative evidence, which may include interviews with experienced market participants, business surveys, past and current market data, as well as more tacit knowledge about the state of the market, will help to narrow the plausible range of π_a for at least some projects.

Evaluating additionality probabilistically may improve DFIs' decision-making, as we do not want DFIs only to invest when they are certain of additionality. Some investments with low values of π_a, which might be thrown out under a binary assessment of additionality, could reveal themselves to have large enough development impacts to compensate for the low anticipated additionality. Moreover, it is often argued that DFIs face a limited supply of viable projects, especially in the countries that stand to gain the most from DFI investment (Kenny, 2019*b*). In this case, only investing when additionality is certain would mean foregoing many worthwhile opportunities to increase investment. The need for investment is so great that we cannot afford missed opportunities, and even though rigorous quantitative evidence of additionality may continue to elude researchers, there is scope for DFIs to improve if they are able to embrace the uncertainty about additionality as a fundamental part of their decision making process.

References

African Development Bank, Asian Development Bank, Asian Infrastructure Investment Bank, European Bank for Reconstruction and Development, European Investment Bank, Inter-American Development Bank, Inter-American Investment Corporation, Islamic Corporation for the Development of the Private Sector, International Finance Corporation, Islamic Development Bank, Multilateral Investment Guarantee Agency, New Development Bank, and World Bank. 2017. "Mobilization of private finance by multilateral development banks: 2016 joint report." http://documents.worldbank.org/curated/en/860721492635844277/2016-joint-report.

African Development Bank, Asian Development Bank, European Bank for Reconstruction and Development, European Investment Bank, Inter-American Development Bank, International Monetary Fund, and World Bank Group. 2015. "From billions to trillions: transforming development finance. Post-2015 financing for development: multilateral development finance." *Development Committee Discussion Note (April 2)*, http://siteresources.worldbank.org/DEVCOMMINT/Documentation/23659446/DC2015-0002(E)FinancingforDevelopment.pdf.

Analysis for Economic Decisions. 2016. "Evaluation of blending: final report." *Evaluation carried out on behalf of the European Commission*, https://ec.europa.eu/europeaid/sites/devco/files/evaluation-blending-volume1_en.pdf.

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

Arellano, Manuel, and Stephen 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.

Attridge, Samantha, and Lars Engen. 2019. "Blended finance in the poorest countries: the need for a better approach." *Overseas Development Institute (ODI) Report*, https://www.odi.org/publications/11303-blended-finance-poorest-countries-need-better-approach.

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

Bazzi, Samuel, and Michael A. Clemens. 2013. "Blunt instruments: avoiding common pitfalls in identifying the causes of economic growth." *American Economic Journal: Macroeconomics*, 5(2): 152–186.

Befani, Barbara, and Gavin Stedman-Bryce. 2017. "Process tracing and Bayesian updating for impact evaluation." *Evaluation*, 23(1): 42–60.

Benn, Julia, Cécile Sangaré, and Tomáš Hos. 2017. "Amounts mobilised from the private sector by official development finance interventions: guarantees, syndicated loans, shares in collective investment vehicles, direct investment in companies and credit lines." *OECD Development Co-operation, Working Paper 36*.

Bischof, Daniel. 2017. "New graphic schemes for Stata: plotplain and plottig." *Stata Journal*, 17(3): 748–759.

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

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

Borusyak, Kirill, Peter Hull, and Xavier Jaravel. 2018. "Quasi-experimental shift-share research designs." *National Bureau of Economic Research (NBER), Working Paper 24997*.

Bourguignon, François, and Mark Sundberg. 2007. "Aid effectiveness - opening the black box." *American Economic Review*, 97(2): 316–321.

Bowsher, Clive G. 2002. "On testing overidentifying restrictions in dynamic panel data models." *Economics Letters*, 77(2): 211–220.

Broccolini, Chiara, Giulia Lotti, Alessandro Maffioli, Andrea F. Presbitero, and Rodolfo Stucchi. 2019. "Mobilization effects of multilateral development banks." *International Monetary Fund, Working Paper WP/19/28*.

Brown, Jessica, Martin Stadelmann, David Wang, Leonardo Boni, Raphaël Jachnik, and Takayoshi Kato. 2015. "Estimating mobilized private finance for adaptation: exploring data and methods." *A Climate Policy Initiative report in collaboration with the OECD*, https://climatepolicyinitiative.org/wp-content/uploads/2015/11/Estimating-mobilized-private-finance-for-adaptation-Exploring-data-and-methods.pdf.

Bun, Maurice J. G., and Monique de Haan. 2010. "Weak instruments and the first stage F-statistic in IV models with a nonscalar error covariance structure." *University of Amsterdam, School of Economics, Department of Quantitative Economics, Discussion Paper 2010/02*.

Bun, Maurice J. G., and Vasilis Sarafidis. 2015. "Dynamic panel data models." In *The Oxford Handbook of Panel Data.* , ed. Badi H. Baltagi, Chapter 3, 76–110. Oxford, United Kingdom:Oxford University Press.

Cáceres Bravo, Mauricio. 2019. "gtools." https://gtools.readthedocs.io/en/latest/index.html.

Card, David. 2001. "Immigrant inflows, native outflows, and the local market impacts of higher immigration." *Journal of Labor Economics*, 19(1): 22–64.

Christian, Paul, and Christopher B. Barrett. 2019. "Spurious regressions and panel IV estimation: revisiting the causes of conflict." *Cornell University, Charles H. Dyson School of Applied Economics, mimeo (May)*, http://barrett.dyson.cornell.edu/files/papers/Spurious%20Regressions%20and%20Panel%20IV%20Estimation%2013%20May%202019%20with%20appendices.pdf.

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

Countdown 2030 Europe. 2018. "Six criteria for donor engagement with the private sector." http://www.countdown2030europe.org/storage/app/media/uploaded-files/C2030E_Six%20Criteria%20for%20Donor%20Engagement%20with%20the%20Private%20Sector_Jan2018.pdf.

Dalberg Global Development Advisors. 2010. "The growing role of the development finance institutions in international development policy." https://www.deginvest.de/DEG-Documents-in-English/Download-Center/EDFI_The-growing-role-of-DFIs-in-International-Development-Policy_2010.pdf.

Dalgaard, Carl-Johan, Henrik Hansen, and Finn Tarp. 2004. "On the empirics of foreign aid and growth." *Economic Journal*, 114(496): F191–F216.

Dawid, Philip, Macartan Humphreys, and Monica Musio. 2019. "Bounding causes of effects with mediators." *arXiv:1907.00399v1*.

Djankov, Simeon, Jose G. Montalvo, and Marta Reynal-Querol. 2008. "The curse of aid." *Journal of Economic Growth*, 13(3): 169–194.

Doucouliagos, Hristos, and Martin Paldam. 2009. "The aid effectiveness literature: the sad results of 40 years of research." *Journal of Economic Surveys*, 23(3): 433–461.

Dreher, Axel, and Sarah Langlotz. 2017. "Aid and growth. New evidence using an excludable instrument." *University of Heidelberg, Department of Economics, Discussion Paper Series No. 635*.

Dreher, Axel, Peter Nunnenkamp, and Rainer Thiele. 2008. "Does aid for education educate children? Evidence from panel data." *World Bank Economic Review*, 22(2): 291–314.

FMO Development Impact Team. 2014. "FMO development impact report 2013/14." *FMO Strategy Department, Development Impact Unit*, https://www.fmo.nl/l/en/library/download/urn:uuid:71ff6ea1-502e-4395-a03b-d706c0890585/fmo+development+impact+report+2013-2014.pdf?format=save_to_disk&ext=.pdf.

Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift. 2019. "Bartik instruments: what, when, why, and how." *Yale School of Managment, mimeo (May)*, https://paulgp.github.io/papers/bartik_gpss.pdf.

Griffith-Jones, Stephany. 2016. "Development banks and their key roles: supporting investment, structural transformation and sustainable development." *Brot für die Welt, Discussion Paper, Analysis 59*, https://www.brot-fuer-die-welt.de/fileadmin/mediapool/2_Downloads/Fachinformationen/Analyse/Analyse59-en-Development_banks_and_their_key_roles.pdf.

Griffiths, Jesse, Matthew Martin, Javier Pereira, and Tim Strawson. 2014. "Financing for development post-2015: improving the contribution of private finance." *Directorate-general for external policies of the Union, Directorate B, Policy Department, study requested by the European Parliament's Committee on Development*, http://eurodad.org/files/pdf/1546190-financing-for-development-post-2015-improving-the-contribution-of-private-finance.pdf.

Gurara, Daniel, Andrea F. Presbitero, and Miguel Sarmiento. 2018. "Borrowing costs and the role of multilateral development banks: evidence from cross-border syndicated bank lending." *International Monetary Fund, Working Paper WP/18/263*.

Hainz, Christa, and Stefanie Kleimeier. 2012. "Political risk, project finance, and the participation of development banks in syndicated lending." *Journal of Financial Intermediation*, 21(2): 287–314.

Hansen, Henrik, and Finn Tarp. 2001. "Aid and growth regressions." *Journal of Development Economics*, 64(2): 547–570.

Hansen, Lars Peter. 1982. "Large sample properties of generalized method of moments estimators." *Econometrica*, 50(4): 1029–1054.

Haščič, Ivan, Miguel Cárdenas Rodríguez, Raphaël Jachnik, Jérôme Silva, and Nick Johnstone. 2015. "Public interventions and private climate finance flows: empirical evidence from renewable energy financing." *OECD Environment Working Papers, No. 80*.

Holtz-Eakin, Douglas, Whitney Newey, and Harvey S. Rosen. 1988. "Estimating vector autoregressions with panel data." *Econometrica*, 56(6): 1371–1395.

Humphreys, Macartan, and Alan M. Jacobs. 2015. "Mixing methods: a Bayesian approach." *American Political Science Review*, 109(4): 653–673.

Independent Evaluation Group. 2008. "Independent evaluation of IFC's development results 2008: IFC's additionality in supporting private sector development." *World Bank Group*, https://openknowledge.worldbank.org/handle/10986/6576.

International Finance Corporation. 2013. "IFC's additionality primer." *mimeo (January 31)*, http://web.worldbank.org/archive/website01555/WEB/IMAGES/ADDITION.DOC?MOD=AJPERES.

Jachnik, Raphaël, Randy Caruso, and Aman Srivastava. 2015. "Estimating mobilised private climate finance: methodological approaches, options and trade-offs." *OECD Environment Working Papers, No. 83*.

Jones, Sam, and Finn Tarp. 2016. "Does foreign aid harm political institutions." *Journal of Development Economics*, 118: 266–281.

Kenny, Charles. 2019*a*. "Five principles for use of aid in subsidies to the private sector." *Center for Global Development (CGD), Policy Paper 157*.

Kenny, Charles. 2019*b*. "Marginal, not transformational: development finance institutions and the sustainable development goals." *Center for Global Development (CGD), Policy Paper 156*.

Kenny, Charles, Jared Kalow, Ben Leo, and Vijaya Ramachandran. 2018. "Comparing five bilateral development finance institutions and the IFC." *Center for Global Development (CGD), Policy Paper 116*.

Kingombe, Christian, Isabella Massa, and Dirk Willem te Velde. 2011. "Comparing development finance institutions: literature review." *Overseas Development Institute (ODI) (20 January)*, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/67635/comparing-DFIs.pdf.

Kleibergen, Frank, and Richard Paap. 2006. "Generalized reduced rank tests using the singular value decomposition." *Journal of Econometrics*, 133(1): 97–126.

Massa, Isabella, Max Mendez-Parra, and Dirk Willem te Velde. 2016. "The macroeconomic effects of development finance institutions in sub-Saharan Africa." *Overseas Development Institute (ODI), Report (December)*, https://www.odi.org/sites/odi.org.uk/files/resource-documents/11182.pdf.

Multilateral Development Banks, and European Development Finance Institutions. 2018. "MDB methodology for private investment mobilization: reference guide." http://documents.worldbank.org/curated/en/495061492543870701/Joint-MDB-reporting-on-private-investment-mobilization-methodology-reference-guide.

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

Olea, José Luis Montiel, and Carolin Pflueger. 2013. "A robust test for weak instruments." *Journal of Business & Economic Statistics*, 31(3): 358–369.

Pereira, Javier. 2015. "Leveraging aid: a literature review on the additionality of using ODA to leverage private investments." *UK Aid Network*, https://web.archive.org/web/20160615074200/http://www.ukan.org.uk:80/wordpress/wp-content/uploads/2015/03/UKAN-Leveraging-Aid-Literature-Review-03.15.pdf.

Rajan, Raghuram G., and Arvind Subramanian. 2008. "Aid and growth: what does the cross-country evidence really show." *Review of Economics and Statistics*, 90(4): 643–665.

Rodrik, Dani, and Arvind Subramanian. 2009. "Why did financial globalization disappoint." *IMF Staff Papers*, 56(1): 1.

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

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

Runde, Daniel F., and Aaron Milner. 2019. "Development finance institutions: plateaued growth, increasing need." *Center for Strategic & International Studies (CSIS), Report (February 13)*.

Sanderson, Eleanor, and Frank Windmeijer. 2016. "A weak instrument F-test in linear IV models with multiple endogenous variables." *Journal of Econometrics*, 190(2): 212–221.

Savoy, Conor M., Paddy Carter, and Alberto Lemma. 2016. "Development finance institutions come of age: policy engagement, impact, and new directions." *A report of the Center for Strategic & International Studies project on prosperity and development and the Overseas Development Institute (ODI)*, https://www.csis.org/analysis/development-finance-institutions-come-age.

Schaffer, Mark E. 2010. "xtivreg2: Stata module to perform extended IV/2SLS, GMM and AC/HAC, LIML and k-class regression for panel data models." http://ideas.repec.org/c/boc/bocode/s456501.html.

Spratt, Stephen, and Lily Ryan Collins. 2012. "Development finance institutions and infrastructure: a systematic review of evidence for development additionality." *A report commissioned by the Private Infrastructure Development Group*, http://www.ids.ac.uk/files/dmfile/2012IDS-PIDGReview2.pdf.

Spratt, Stephen, Chris Barnett, Peter O'Flynn, Vishal Gadhavi, Jean-Pierre Tranchant, Amrita Saha, Mel Punton, and Katie Spafford. 2019. "Inception report: study of CDC's mobilisation of private investment." *Submitted by Itad, in association with Institute of Development Studies, Open Capital Advisors and Ipsos MORI*, https://www.itad.com/reports/inception-report-study-of-cdcs-mobilisation-of-private-investment/.

Stock, James H., and Motohiro Yogo. 2005. "Testing for weak instruments in linear IV regression." In *Identification and inference for econometric models: essays in honor of Thomas Rothenberg.* , ed. Donald W. K. Andrews and James H. Stock, 80–108. Cambridge, United Kingdom:Cambridge University Press.

Temple, Jonathan, and Nicolas Van de Sijpe. 2017. "Foreign aid and domestic absorption." *Journal of International Economics*, 108: 431–443.

te Velde, Dirk Willem. 2011. "The role of development finance institutions in tackling global challenges." *Overseas Development Institute (ODI)*, https://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/7310.pdf.

The Association of European Development Finance Institutions. 2016. "Flagship report 2016: European development finance for private sector projects in low and middle income countries." https://www.edfi.eu/news/flagship-report-2016/.

The European Union blending and external cooperation (EUBEC) platform. 2013. "Enhancement of blending activities. Initial topic: measuring results, monitoring and reporting." *Report by technical group 2 (TG2) to the policy group of the European Union blending and external cooperation (EUBEC) platform (June)*, https://ec.europa.eu/transparency/regexpert/index.cfm?do=groupDetail.groupDetailDoc&id=19079&no=6.

Wansbeek, Tom. 2012. "On the remarkable success of the Arellano-Bond estimator." *AENORM*, 20(77): 15–20.

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

Windmeijer, Frank. 2018. "Testing over- and underidentification in linear models, with applications to dynamic panel data and asset-pricing models." *University of Bristol, Department of Economics, Discussion Paper 18/696.*

Xu, Jiajun, Xiaomeng Ren, and Xinyue Wu. 2019. "Mapping development finance institutions worldwide: definitions, rationales, and varieties." *Peking University, Institute of New Structural Economics, Development Financing Research Report No. 1.*

Appendix A Additional results and discussion

A.1 OLS and fixed effects estimation with an alternative project selection mechanism

In the paper we show results for OLS and FE estimation when DFIs randomly select projects from their eligible set until the DFI budget runs out. Here, we consider an alternative where the DFI sector, from its eligible set, first picks the projects with the worst project characteristics. This alternative selection mechanism could be interpreted as DFIs attempting to (be seen to) fulfill a mandate to do deals in difficult markets by prioritising those projects that superficially look like ones the private sector would avoid, for instance because these projects are located in countries with generally weaker investment climates.[30] For example, DFIs might prefer to invest in a project in the DRC over one in Brazil if both projects offer the same risk-adjusted expected return. Table A.1 shows results for our default DGP; the only difference with Table 1 is the project selection mechanism. As rejection rates for the true nulls now sometimes deviate from 100%, we add these to the table as well.

Table A.1: OLS and fixed effects results with an alternative DFI project selection mechanism

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
β	0	0	0	0	1	1	1	1
pc	excl.	$\sigma_m^2 = 1$	$\sigma_m^2 = 0.5$	$\sigma_m^2 = 0$	excl.	$\sigma_m^2 = 1$	$\sigma_m^2 = 0.5$	$\sigma_m^2 = 0$
Mean $\widehat{\beta}_{OLS}$	-0.76	-0.10	0.08	0.16	-1.57	-0.62	0.18	0.91
Std. dev.	0.80	0.21	0.05	0.06	0.47	0.17	0.08	0.05
% reject $\beta \leqslant 0$	11.2	16.4	64.5	100				
% reject $\beta \geqslant 1$					100	100	100	86.1
Mean $\widehat{\beta}_{FE}$	-0.50	-0.06	0.11	0.20	-1.37	-0.60	0.13	0.90
Std. dev.	0.62	0.22	0.06	0.06	0.40	0.18	0.08	0.06
% reject $\beta \leqslant 0$	12.7	21.2	70.9	100				
% reject $\beta \geqslant 1$					100	100	100	88.4

Note: mean value and standard deviation of OLS and fixed effects estimates of β, based on 1000 replications of our DGP. % reject $\beta \leqslant 0$ and % reject $\beta \geqslant 1$ are the percentages of replications in which the null of zero additionality and the null of full additionality are rejected at a 5% significance level, respectively. DFIs first pick the projects with the worst project characteristics from their eligible set. pc indicates when $\widetilde{pc}_{it}$ is excluded ('excl.') or, when it is included, how much measurement error has been added to it.

Consider first the zero additionality case (columns 1-4). As DFIs and private investors are interested in the same pool of projects (whose returns exceed the common lower bound), the rationale for an upward bias outlined in the paper still holds. However, by prioritising, within this eligible set, the projects with the weakest project characteristics, some DFI investment will flow to country-periods with lower overall investment. This is because, in the global set of projects with sufficient returns, the projects with the lowest values for pc_{pit} are more likely to be drawn from

[30] In our DGP, pc_{pit} is a less noisy measure of a country's type than er_{pit}, since the latter contains an additional project-specific component (e_{pit}).

country-periods with low average returns and, hence, low private investment. This alternative selection mechanism thus pushes against the upward bias. On balance, we may even end up with a net downward bias. This is the case, for instance, in column 1, where project characteristics are not controlled for. Rejection rates for zero additionality fall correspondingly compared to the random project selection mechanism. The change in the size and even sign of the bias when we vary how DFIs select projects further illustrates how the estimated degree of additionality can reflect the particular way in which DFIs attempt to fulfill their mandate.

Under full additionality (columns 5-8), we find a similar downward bias as for random project selection. The coefficients in this scenario are mostly determined by the two types of investors targeting very different returns. Conditional on this, it matters less whether the DFI sector chooses projects at random from its eligible set, or whether it first picks the projects with the worst observable characteristics.[31]

A.2 Supply-push IV

A.2.1 Additional discussion of endogenous DFI budgets

In our DGP, we assume that the global DFI budget is exogenous to give us a benchmark in which the supply-push instrument is valid. As a result, the endogeneity of the instrument comes only from endogenous reactions of individual DFI budgets combined with DFI-specific preferences for some countries over others. In practice, however, the shared preference of DFIs for high (under zero additionality) or low (under full additionality) returns could also make the IV estimator inconsistent if the global DFI budget is a function of the number of projects DFIs are interested in. An example can clarify. Suppose, in the zero additionality version of our DGP, that a few countries experience positive shocks to u_{it} that increase the amount of DFI investment they receive. As DFIs respond to the increase in investment opportunities, the global DFI budget rises (D_{dt} increases for most DFIs). If the countries that experience the positive shocks are also the ones with large initial shares s^d_{i0}, the instrument $dfiIV_{it}$ for these countries will increase in tandem with dfi_{it}. As a result, $dfiIV_{it}$ will be positively correlated with u_{it}, again resulting in upward bias.

This also makes clear that, when DFI budgets respond endogenously to the number of in-

[31]The alternative selection mechanism does not always have a larger downward bias than the random selection mechanism. The OLS bias, for instance, is given by $1\frac{\mathrm{Cov}\left(dfi^{\perp}_{it}, \#er>2^{\perp}_{it}\right)}{\mathrm{Var}\left(dfi^{\perp}_{it}\right)}$, where $\perp$ indicates that time dummies and, when included, $\widetilde{pc}_{it}$ have first been partialled out. $\#er > 2_{it}$ is the number of projects with expected returns over 2, which enters the true model with a coefficient of 1, and whose inclusion would completely remove the bias in the estimation of β. The alternative selection mechanism tends to concentrate DFI investment in a smaller number of (low-type) countries, in most cases increasing $\mathrm{Var}\left(dfi^{\perp}_{it}\right)$ and making $\mathrm{Cov}\left(dfi^{\perp}_{it}, \#er > 2^{\perp}_{it}\right)$ more negative. The change in the bias from switching from the random to the alternative selection mechanism then depends on whether the absolute value of the covariance or variance rises the most proportionally.

vestment opportunities, a likely tradeoff between instrument strength and validity surfaces. As explained in the previous paragraph, if the shocks that attract more DFI investment occur in countries with large initial shares s^d_{i0}, the instrument becomes invalid. If, in contrast, the countries experiencing these shocks are the ones with low s^d_{i0}, then the instrument will weaken as it will fail to track the increase in actual DFI investment received by these countries.

A.2.2 Leave-one-out version of the instrument

To deal with feedback from an individual unit to the aggregate shifter, researchers often use a leave-one-out version of their shift-share instrument (see, for instance, Goldsmith-Pinkham, Sorkin and Swift, 2019, for a discussion). For our supply-push instrument, this means replacing D_{dt} in equation (8) with $(D_{dt} - dfi_{it})$, which is the amount of DFI d's total budget in t allocated to countries other than country i. In our set-up, this does not solve the problem, however. In its first seven columns, Table A.2 repeats the results from Table 2 in the main text. Recall that these results show how the endogeneity of DFIs' budgets, introduced by setting a positive value for ϕ, generates an upward bias in the supply-push IV estimator. Column 8 establishes that applying the leave-one-out version of the instrument does not solve the problem of endogenous budgets. This column shows that the leave-one-out IV estimator yields a downward bias when it is applied to the configuration from column 5 (with $\phi = 1$). Our global DFI budget is exogenous, so it is not affected by shocks to countries' investment returns. As a result, when country i experiences a positive shock and receives more DFI investment, less DFI investment is available for other countries, so that $D_{dt} - dfi_{it}$ for country i might fall for many DFIs. When this happens it can lead to a negative correlation between u_{it} and the leave-one-out version of the instrument, as is the case in column 8.

A.2.3 Results with an upward trending budget

The situation with an upward trending DFI budget, considered in Table A.3, is slightly more complicated than the case with a downward trending budget discussed in the paper.[32] The differential evolution of total investment is similar: negative for initial high-type countries and positive for countries that start out as low-type. In contrast, the upward trend in the budget leads to a more rapid rise over time in the instrument for initial high-type countries than for initial low-type countries. The opposing differential changes in total investment and the instrument result in a negative reduced form coefficient in all but one replication in columns 1 and 2. In

[32]The upward drift we consider is smaller in size than the downward drift in the paper. This is to make sure that the global DFI budget never exceeds the number of projects with an expected return over 2.

Table A.2: Supply-push IV results with zero additionality ($\beta = 0$): leave-one-out version of the instrument

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Δ types	default	fewer	1period	fixed	1period	1period	1period	1period
ϕ	0	0	0	0	1	2	2	1
pc	excl.	excl.	excl.	excl.	excl.	excl.	$\sigma_m^2 = 0.5$	excl.
σ_{db}	0.15	0.15	0.15	0.15	0.15	0.15	0.05	0.15
Instrument	standard	standard	standard	standard	standard	standard	standard	leave1out
Med. $\widehat{\beta}_{IV}$	4.56	0.53	0.00	0.00	0.12	0.26	0.39	-0.52
Std. dev.	84.72	19.65	0.97	0.09	0.63	0.54	0.53	53.66
% reject $\beta \leqslant 0$	68.6	43.3	8.2	4.7	15.9	27.4	29.7	4.1
Med. F	4.77	23.6	37.1	166	43.7	49.4	20.5	17.3
% reject underid.	53.4	72.9	97	100	98.8	99.7	97.4	77

Note: median value and standard deviation of IV estimates of β, based on 1000 replications of our DGP. % reject $\beta \leqslant 0$ is the percentage of replications in which the null of zero additionality is rejected at a 5% significance level. The final two rows show the median cluster-robust first-stage F-statistic, and the percentage of replications that reject underidentification at a 5% significance level. DFIs randomly select projects from their eligible set. $\Delta types$ states how often types change over time: as given by transition matrix (3) ('default'); reduced probabilities of transitions as in transition matrix (9) ('fewer'); transitions occur for one period only ('1period'); types are fixed ('fixed'). *pc* indicates when $\widetilde{pc}_{it}$ is excluded ('excl.') or, when it is included, how much measurement error has been added to it. Instrument denotes whether the standard version of the instrument is used or the leave-one-out ('leave1out') version.

column 1, with default type transitions, the majority of first stage coefficients are also negative, producing a positive median IV estimate. The first stage coefficient is so often negative because initial low-type countries tend to see larger increases in DFI investment than countries that start out as high-type, which is the opposite pattern as that for the instrument. The reason for this is that, while both sets of countries benefit from the increasing DFI budget, when type transitions are frequent, low-type countries tend to shift up type over time, further increasing their DFI investment, while countries that start as high-type tend to shift down in type, which lowers their DFI investment.

Table A.3: Supply-push IV results with zero additionality ($\beta = 0$): upward trend in the DFI budget

	(1)	(2)	(3)	(4)	(5)
drift	0.05	0.05	0.05	0.05	0.05
Δ types	default	fewer	fixed	default	default
i.trend	no	no	no	yes	no
pc	excl.	excl.	excl.	excl.	$\sigma_m^2 = 0.5$
Med. $\widehat{\beta}_{IV}$	3.96	-2.66	0.00	0.86	-1.98
Std. dev.	7.10	693.04	0.03	5376.09	144.65
% reject $\beta \leqslant 0$	90.1	7.5	6	31.5	1.7
Med. F	5.74	.545	422	1.67	1.02
% reject underid.	63	9.7	100	12.2	14.3
Med. FS	-0.49	0.05	0.97	1.20	0.08
Med. RF	-1.99	-1.01	0.00	0.95	-0.29

Note: see Table A.2. This table also shows median first stage ('FS') and reduced form ('RF') estimates. For all results in this table $\sigma_{db} = 0$. i.trend indicates in which columns time dummies have been replaced by country-specific trends.

When types change less frequently (column 2), the main difference is that DFI investment tends to rise more quickly for initial high-type countries than for countries starting out as low-type, because the former benefit more from the expansion of the DFI budget, and because the reduction in type changes takes away most of the negative (positive) pressure on DFI investment for initial high-type (low-type) countries. As a result, the differential changes in DFI investment more often match those in the instrument, turning more of the first stage estimates positive and more of the IV estimates negative, resulting in a downward bias in column 2. In column 3, with fixed types, the bias again disappears as the median reduced form coefficient becomes zero.

Column 4 shows that, for the default transition mechanism, replacing time dummies by country-specific trends is insufficient to fully remove the bias. Likewise, controlling for $\widetilde{pc}_{it}$ (with $\sigma_m^2 = 0.5$) in column 5 reduces the bias but does not eliminate it. For an upward trending global DFI budget, controlling for $\widetilde{pc}_{it}$ also turns more of the estimated first stage coefficients positive, so that the median bias changes sign compared to the case without $\widetilde{pc}_{it}$.

A.2.4 Results under full additionality

The first four columns in Table A.4 show how making types more persistent increases instrument strength and reduces bias. As was the case for zero additionality, restricting transitions to a single period in column 3 is sufficient to remove all bias. Proceeding with this single period transition case, columns 5 and 6 show how setting $\phi > 0$ introduces bias that rises with the value of ϕ. As in the main text, the reason for this is the endogenous reaction of overall DFI budgets to shocks to expected returns in countries that a DFI has a strong preference for. Column 7 repeats the example from the main text to show that, even with $\widetilde{pc}_{it}$ included (with a small amount of measurement error: $\sigma_m^2 = 0.5$) a substantial bias may remain. Column 8 applies the leave-one-out IV estimator to the configuration from column 5 (with $\phi = 1$), showing a small upward bias. When a country experiences a negative shock and receives more DFI investment, less DFI investment is available for other countries, so that $D_{dt} - dfi_{it}$ might fall for many DFIs. This can lead to a positive correlation between u_{it} and the leave-one-out version of the instrument, and an upward bias, as is the case in column 8.

Tables A.5 and A.6 explore how the combination of trends in the global DFI budget and countries changing types can lead to bias. Table A.5 considers a negative trend. In column 1 we consider default type transitions, while column 2 examines what happens with reduced probabilities of type changes. In both cases, the median estimate is negative, suggesting a large downward bias. Similar mechanisms as discussed in the main text are at play. With a downward trending

Table A.4: Supply-push IV results with full additionality ($\beta = 1$)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Δ types	default	fewer	1period	fixed	1period	1period	1period	1period
ϕ	0	0	0	0	1	2	2	1
pc	excl.	excl.	excl.	excl.	excl.	excl.	$\sigma_m^2 = 0.5$	excl.
σ_{db}	0.15	0.15	0.15	0.15	0.15	0.15	0.05	0.15
Instrument	standard	standard	standard	standard	standard	standard	standard	leave1out
Med. $\widehat{\beta}_{IV}$	-3.48	0.47	1.00	1.00	0.74	0.53	0.49	1.04
Std. dev.	209.64	106.80	8.40	0.50	1.36	0.96	0.57	998.41
% reject $\beta \geqslant 1$	57.3	36.5	8.2	5.5	15	23.8	29.3	14.3
Med. F	3.53	11.7	14.8	45.6	22.4	29.4	20.5	2.91
% reject underid.	46	72.1	85.1	98.5	97.5	99.4	99.5	42.2

Note: see Table A.2. % reject $\beta \geqslant 1$ is the percentage of replications in which the null of full additionality is rejected at a 5% significance level.

budget, under full additionality DFI investment falls most rapidly for countries that start out as low-type, and the same goes for the instrument, generating a positive first stage relationship. Total investment, however, falls most rapidly for countries that are initially high-type, producing negative reduced form estimates. The combination of positive first stage estimates and negative reduced form estimates yields negative IV estimates. Keeping types fixed again eliminates most of the bias (column 3). Controlling for country-specific trends (column 4) or $\widetilde{pc}_{it}$ (with $\sigma_m^2 = 0.5$, column 5) does not get rid of the bias.

Table A.5: Supply-push IV results with full additionality ($\beta = 1$): downward trend in the DFI budget

	(1)	(2)	(3)	(4)	(5)
drift	-0.1	-0.1	-0.1	-0.1	-0.1
Δ types	default	fewer	fixed	default	default
i.trend	no	no	no	yes	no
pc	excl.	excl.	excl.	excl.	$\sigma_m^2 = 0.5$
Med. $\widehat{\beta}_{IV}$	-8.29	-4.43	1.01	-3.42	-1.23
Std. dev.	2.62	2.00	0.29	3.57	1.57
% reject $\beta \geqslant 1$	96	91.3	5.9	42.5	59.8
Med. F	21.1	38.3	44.4	9.56	12.7
% reject underid.	98	99.8	100	100	90.3
Med. FS	0.86	0.82	0.73	1.19	0.55
Med. RF	-7.15	-3.52	0.74	-3.99	-0.67

Note: see Table A.2. % reject $\beta \geqslant 1$ is the percentage of replications in which the null of full additionality is rejected at a 5% significance level. This table also shows median first stage ('FS') and reduced form ('RF') estimates. For all results in this table $\sigma_{db} = 0$. i.trend indicates in which columns time dummies have been replaced by country-specific trends.

Likewise, in Table A.6, the biases for an upward trending budget are of the opposite sign as for the zero additionality case discussed in the main text. With fixed types (column 3), there is no differential change in total investment by initial type, and the median IV estimate equals the true β.

Table A.6: Supply-push IV results with full additionality ($\beta = 1$): upward trend in the DFI budget

	(1)	(2)	(3)	(4)	(5)
drift	0.05	0.05	0.05	0.05	0.05
Δ types	default	fewer	fixed	default	default
i.trend	no	no	no	yes	no
pc	excl.	excl.	excl.	excl.	$\sigma_m^2 = 0.5$
Med. $\widehat{\beta}_{IV}$	-3.30	4.87	1.00	-2.07	1.73
Std. dev.	23.10	269.65	0.09	19.17	702.10
% reject $\beta \geqslant 1$	71.4	1	6.7	24.2	3.2
Med. F	2.99	1.41	92.9	1.04	.492
% reject underid.	37.9	22.9	100	1.6	5.6
Med. FS	-0.63	0.28	0.80	1.07	0.06
Med. RF	2.19	1.93	0.80	-2.21	0.47

Note: see Table A.5.

A.3 System GMM

A.3.1 Additional experiments

In Table A.7 we report four additional experiments to give further insight into the performance of the system GMM estimator. For each experiment we first report results from a regression that excludes $\widetilde{pc}_{it}$, then from a regression that includes $\widetilde{pc}_{it}$ with the smallest amount of measurement error added to it ($\sigma_m^2 = 0.5$). We now also show median difference ($\widehat{\beta}_{diffGMM}$) and levels ($\widehat{\beta}_{levGMM}$) GMM estimates, obtained by separately estimating the differenced and levels equations, as these will prove useful to interpret the findings of one of the experiments in this table.

We first demonstrate that our findings do not depend on the particular formulation of the DFI sector's budget in the default version of our DGP. In columns 1-2, we remove the stochastic element in the DFI budget by setting $\sigma_{db} = 0$, and replace it with a deterministic upward trend (*drift* $= 0.05$). This mimics the rise in DFI investments seen over the past two decades or so (Runde and Milner, 2019). The upward trending budget will also prove useful for one of the later experiments. As was the case before, system GMM suffers from an upward bias.

In our default set-up countries of different types have very different probabilities of receiving DFI investment. Given the crucial role played by changes in types for the violation of the moment conditions in (11), it is important to see whether the bias in system GMM can be removed by narrowing the gap in average expected returns between types. By making types more similar, the correlation between lagged DFI investment and contemporaneous changes in the number of high return projects should weaken; and likewise for the correlation between lagged changes in DFI investment and the current number of projects with sufficient expected returns. We first change mean pc_{pit} for the three types from $\mu_c = [0, 2, 4]$ to $\mu_c = [1, 2, 3]$, then also consider a case where

Table A.7: System GMM results with zero additionality ($\beta = 0$): additional experiments

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Experiment description	Upward trend in DFI budget		μ_c closer together		μ_c the same		Upward trend in DFI budget, fixed types	
pc	excl.	$\sigma_m^2 = 0.5$	excl.	$\sigma_m^2 = 0.5$	excl.	$\sigma_m^2 = 0.5$	excl.	$\sigma_m^2 = 0.5$
σ_{db}	0	0	0.15	0.15	0.15	0.15	0	0
drift	0.05	0.05	0	0	0	0	0.05	0.05
μ_c	$[0,2,4]$	$[0,2,4]$	$[1,2,3]$	$[1,2,3]$	$[2,2,2]$	$[2,2,2]$	$[0,2,4]$	$[0,2,4]$
Δ types	default	default	default	default	default	default	fixed	fixed
Med. $\widehat{\beta}_{sysGMM}$	2.58	0.07	3.53	0.59	0.53	0.48	1.17	0.02
Std. dev.	0.26	0.35	2.06	1.76	3.21	1.53	0.05	0.13
% reject $\beta \leqslant 0$	100	10.5	98.2	20.9	3.5	5.1	100	8.7
Hansen % pass	.2	91.1	64.6	89.2	94.4	97.5	0	89.5
Med. cond. F	40.4	20.9	14.9	5.44	1.44	1.88	40.3	33.8
% reject underid.	100	97.5	91.7	35.5	3.5	2.5	100	100
Med. $\widehat{\beta}_{diffGMM}$	3.69	0.73	4.30	1.30	0.54	0.64	0.00	0.00
Med. $\widehat{\beta}_{levGMM}$	2.92	0.46	4.22	1.32	0.65	0.60	2.00	1.19

Note: median value and standard deviation of system GMM estimates of β, based on 1000 replications of our DGP. % reject $\beta \leqslant 0$ is the percentage of replications in which the null of zero additionality is rejected at a 5% significance level. Hansen % pass is the percentage of replications that do not reject Hansen's overidentifying restrictions test at a 10% significance level. The next two rows report the median cluster-robust conditional F-statistic, and the percentage of replications that reject underidentification at a 5% significance level. The final two rows report the median value of difference and levels GMM estimates of β. DFIs randomly select projects from their eligible set. *pc* indicates when $\widetilde{pc}_{it}$ is excluded ('excl.') or, when it is included, how much measurement error has been added to it. Types change over time ('Δ types') according to the transition matrix (3) ('default') or are time-invariant ('fixed'). For ease of interpretation, the top row shows a brief description of the experiment considered.

average project characteristics are equal to 2 regardless of a country's type ($\mu_c = [2,2,2]$).

Looking at the model without $\widetilde{pc}_{it}$, we can see changes in results that are consistent with a weakening of violations of the moment conditions: bringing average returns closer together reduces the Hansen test's rejection rate from around 67% in column 1 of Table 4 to 35.4% in column 3 of Table A.7, while setting $\mu_c = 2$ for all types further reduces the rejection rate to around 5% (column 5 in Table A.7). Nonetheless, compared to the default set-up in Table 4 there is almost no reduction in bias from bringing average returns closer together, and even when all types are the same, so that the moment conditions in (11) should be satisfied, the GMM estimators still display bias. The reason for this is that a movement of Corr $\left(dfi_{i,t-2}, \Delta u_{it}\right)$ and Corr $\left(\Delta dfi_{i,t-1}, w_i + u_{it}\right)$ towards zero is matched by a weakening of Corr $\left(dfi_{i,t-2}, \Delta dfi_{it}\right)$ and Corr $\left(\Delta dfi_{i,t-1}, dfi_{it}\right)$: as lagged levels and differences of DFI investment are no longer correlated with current changes in, and levels of, the number of projects with sufficient returns, they also lose their ability to predict contemporaneous changes in, and levels of, DFI investment. When $\mu_c = 2$ for all types, the instruments have no strength left, and underidentification is rejected in fewer than 5% of the replications. The upshot is bias due to weak instruments. When we add $\widetilde{pc}_{it}$ to the model (column 6), the bias barely changes, because the additional instruments based on lagged levels and differences of $\widetilde{pc}_{it}$ are also uninformative.

These results suggest that changes in types create a tradeoff between instrument strength and validity in system GMM, where larger gaps in average expected returns between countries of different types strengthen instruments but at the same time exacerbate moment violations. In the final experiment in Table A.7, we examine how a trend in the DFI sector's budget can relax this tradeoff. In this experiment we revert to default values for μ_c but make types time-invariant. To create some instrument strength, we rely on an upward trending budget.[33] Again, we start by considering the model without $\widetilde{pc}_{it}$ (column 7).

In the differenced equation, fixing types gets rid of moment violations: there is now no reason for $dfi_{i,t-2}$ to be systematically correlated with Δu_{it}. If the DFI sector's budget were flat in expectation over the sample period, there would also be no reason for $dfi_{i,t-2}$ to be correlated with Δdfi_{it}. A trend in the DFI budget, however, can generate instrument strength in the differenced equation even with fixed types. This is because high-type countries, who have, on average, more investable projects, benefit more from an expanding DFI budget than low-type countries, who have few projects DFIs are willing to invest in. As a result, changes in DFI investment will be more positive for high-type countries, who also have higher lagged levels of DFI investment. This results in a situation where $\text{Corr}\left(dfi_{i,t-2}, \Delta dfi_{it}\right) > 0$ even though $\text{Corr}\left(dfi_{i,t-2}, \Delta u_{it}\right) = 0$. In column 7, even without controlling for $\widetilde{pc}_{it}$, the median bias in difference GMM disappears.

Interestingly, the same is not the case for levels GMM. Even with fixed types, the greater increases in DFI investment for high-type countries when the DFI budget trends upward imply a positive correlation between $\Delta dfi_{i,t-1}$ and w_i, so that a moment condition for the levels equation is not satisfied. As a result, both levels and system GMM are biased. This also explains why the Hansen test for system GMM rejects in every single replication, as one set of moment conditions (those for the differenced equation) holds while a different set does not. Adding $\widetilde{pc}_{it}$ again reduces this bias, but given the obvious violation of some of the moment conditions system GMM relies on, there is little guarantee that the upward bias in system GMM will always be as low as it is in column 8, even with fixed types and a positive trend in the budget.

A.3.2 Results under full additionality

For completeness, we briefly discuss system GMM results under full additionality, where OLS and FE underestimate the true $\beta = 1$. The results are the mirror image of those reported for zero additionality.

[33] A similar narrative can be developed for a downward trending budget.

Table A.8: System GMM results with full additionality ($\beta = 1$)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
LDV	no	no	no	no	yes	yes	yes	yes
pc	excl.	$\sigma_m^2 = 1$	$\sigma_m^2 = 0.5$	$\sigma_m^2 = 0$	excl.	$\sigma_m^2 = 1$	$\sigma_m^2 = 0.5$	$\sigma_m^2 = 0$
Med. $\widehat{\beta}_{sysGMM}$	-3.34	0.64	0.93	0.97	-1.46	0.07	0.67	0.93
Std. dev.	2.12	2.20	1.07	0.40	1.50	1.46	0.92	0.41
% reject $\beta \geqslant 1$	99.5	18.4	9.6	7.4	96.1	37.7	17.1	8.3
Hansen % pass	59.1	87	88.6	91.1	51.3	87	90	90.9
Med. cond. F	27.9	9.63	14.3	14.9	24.2	6.28	7.46	10.2
% reject underid.	97.8	59.2	71.7	74.5	92	30.2	38.1	54.4

Note: see Table A.7. % reject $\beta \geqslant 1$ is the percentage of replications in which the null of full additionality is rejected at a 5% significance level. LDV indicates whether a lagged dependent variable is included.

The first column of Table A.8 shows results for default values of the parameters in our DGP, for a regression without $\widetilde{pc}_{it}$. Underidentification is rejected in almost every replication, but the Hansen test also often rejects and system GMM shows a large downward bias, because the moment conditions in (11) are not satisfied.

Under full additionality, a low-type country receives more DFI investment than a high-type country, and, through type changes, is more likely to experience an increase in the number of projects with high expected returns in the future, which would decrease DFI investment. The consequence is that Corr $\left(dfi_{i,t-2}, \Delta dfi_{it}\right) < 0$ and Corr $\left(dfi_{i,t-2}, \Delta u_{it}\right) > 0$, leading to a downward bias.

For the levels equations, if a high-type country moves down types in $t-1$, its DFI investment increases, and, since types are persistent, it is also likely to end up with fewer projects with high expected returns in period t, implying that Corr $\left(\Delta dfi_{i,t-1}, dfi_{it}\right) > 0$ but also that Corr $\left(\Delta dfi_{i,t-1}, w_i + u_{it}\right) < 0$. Even when types are time-invariant, trends in the DFI sector's budget can again contribute to violations of the moment conditions in the levels equation, as we discuss below.

Including $\widetilde{pc}_{it}$, especially without measurement error, reduces bias (see columns 2-4 in Table A.8) as it partially controls for the number of projects with expected returns over 2, weakening the correlations between instruments and error terms. As was the case under zero additionality, however, there is nothing inherent about system GMM that removes the bias in the estimation of β. This conclusion is unaltered when we include a lagged dependent variable in the final four columns of Table A.8.

The downward bias is still apparent in Table A.9 when we consider an upward trending budget for the DFI sector (see especially column 1, without $\widetilde{pc}_{it}$). Narrowing the gap in average expected returns between countries of different types (columns 3-4), or even setting $\mu_c = 2$

Table A.9: System GMM results with full additionality ($\beta = 1$): additional experiments

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Experiment description	Upward trend in DFI budget		μ_c closer together		μ_c the same		Upward trend in DFI budget, fixed types	
pc	excl.	$\sigma_m^2 = 0.5$	excl.	$\sigma_m^2 = 0.5$	excl.	$\sigma_m^2 = 0.5$	excl.	$\sigma_m^2 = 0.5$
σ_{db}	0	0	0.15	0.15	0.15	0.15	0	0
$drift$	0.05	0.05	0	0	0	0	0.05	0.05
μ_c	$[0, 2, 4]$	$[0, 2, 4]$	$[1, 2, 3]$	$[1, 2, 3]$	$[2, 2, 2]$	$[2, 2, 2]$	$[0, 2, 4]$	$[0, 2, 4]$
Δ types	default	default	default	default	default	default	fixed	fixed
Med. $\widehat{\beta}_{sysGMM}$	-1.17	0.97	-3.03	0.33	0.40	0.43	-1.03	0.87
Std. dev.	0.24	0.18	5.19	1.90	2.71	1.98	0.31	0.39
% reject $\beta \geqslant 1$	100	8.3	85.4	16.9	6.4	5.6	100	12.7
Hansen % pass	.4	90.2	67.8	91	94	96.9	49.5	87.6
Med. cond. F	52.4	38	9.13	4.83	1.53	1.92	17.4	12.8
% reject underid.	100	100	66.6	29.9	5.1	1.6	100	84.8
Med. $\widehat{\beta}_{diffGMM}$	-2.82	0.99	-4.48	-0.20	0.33	0.41	0.95	0.96
Med. $\widehat{\beta}_{levGMM}$	-1.71	0.99	-3.84	-0.40	0.42	0.43	-2.81	-0.03

Note: see Table A.7. % reject $\beta \geqslant 1$ is the percentage of replications in which the null of full additionality is rejected at a 5% significance level.

regardless of type (columns 5-6), again does not remove the bias in system GMM estimation. Fixing types and relying on an upward trending budget for instrument strength (columns 7-8) yields little bias for difference GMM even without controlling for $\widetilde{pc}_{it}$, but the same is not true for levels or system GMM. A plausible reason for why there is still a small amount of bias left in the difference GMM case, in contrast to the zero additionality case considered earlier, is that the instruments are less strong here (the median conditional F-statistic is 17.4, compared to 40.3 in the zero additionality case).