Online Appendix for "The Elusive Quest for Additionality"

December 2, 2020

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 prioritizing 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.[1] 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.

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 prioritizing, 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 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

[1] 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}).

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: this table shows mean values and standard deviations 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.

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.[2]

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

[2]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 $1\frac{\mathrm{Cov}\left(dfi_{it}^{\perp}, \#er>2_{it}^{\perp}\right)}{\mathrm{Var}\left(dfi_{it}^{\perp}\right)}$, 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 two, which enters the true model with a coefficient of one, 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 $\mathrm{Var}\left(dfi_{it}^{\perp}\right)$ and making $\mathrm{Cov}\left(dfi_{it}^{\perp}, \#er > 2_{it}^{\perp}\right)$ 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.

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_{i0}^d, 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 investment 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_{i0}^d, the instrument becomes invalid. If, in contrast, the countries experiencing these shocks are the ones with low s_{i0}^d, 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. For our supply-push instrument, this means replacing D_{dt} in equation (8) with $\left(D_{dt} - dfi_{it}^d\right)$, 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 $D_{dt} - dfi_{it}^d$ 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.

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: this table shows median values and standard deviations 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. Δ*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.

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.[3] 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 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.

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-

[3]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 two.

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.	$\sigma^2_m = 0.5$
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.

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 $\sigma^2_m = 0.5$) 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 examine how changing the type transition mechanism affects instrument strength and bias under full additionality. 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: $\sigma^2_m = 0.5$) 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 $D_{dt} - dfi_{it}^{d}$ 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.

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.

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 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.

Likewise, in Table A.6, the biases for an upward trending budget are of the opposite sign as for the zero additionality case discussed earlier. 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.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.

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 Main results

Difference GMM (Arellano and Bond, 1991; Holtz-Eakin et al., 1988) starts by differencing equation (1) to remove w_i:

$$\Delta I_{it} = \beta \Delta dfi_{it} + \gamma \Delta \widetilde{pc}_{it} + \Delta \delta_t + \Delta u_{it} \tag{A.1}$$

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

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, 2009b), we use only a single lagged level of each variable as an instrument for the differenced equations. This yields the following population moment conditions:[4]

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

To further counter overfitting, we also collapse the instrument matrix (Roodman, 2009a). 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 (A.2) 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 A.7 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.[5]

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.

Table A.7 shows results for the same default version of our DGP that was used in the first four

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

[5]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 A.7: 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: this table shows median values and standard deviations 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.

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 (A.2) 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.[6]

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 number of projects with high expected returns in period t, implying that $\text{Corr}\left(\Delta dfi_{i,t-1}, dfi_{it}\right) > 0$

[6] 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.

but also that 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.

As was the case for the other estimators, including $\widetilde{pc}_{it}$, especially without measurement error, reduces bias (columns 2-4 in Table A.7). $\widetilde{pc}_{it}$ partially controls for the number of projects with sufficient expected returns, weakening the correlations between instruments and error terms. Table A.7 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 A.7 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 $\mathrm{E}\left[I_{i,t-2}\Delta u_{it}\right] = 0$ and $\mathrm{E}\left[\Delta I_{i,t-1}\left(w_i + u_{it}\right)\right] = 0$. The main change from including $I_{i,t-1}$ is lower bias in the model without $\widetilde{pc}_{it}$.

A.3.2 Additional results

In Table A.8 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

Table A.8: 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.	$\sigma^2_m = 0.5$	excl.	$\sigma^2_m = 0.5$	excl.	$\sigma^2_m = 0.5$	excl.	$\sigma^2_m = 0.5$
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: see Table A.7. The final two rows report the median value of difference and levels GMM estimates of β. 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.

conditions in (A.2), 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 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 A.7 to 35.4% in column 3 of Table A.8, while setting $\mu_c = 2$ for all types further reduces the rejection rate to around 5% (column 5 in Table A.8). Nonetheless, compared to the default set-up in Table A.7 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 (A.2) 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.8, 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.[7] 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.

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

A.3.3 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.

Table A.9: 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.8. % 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.9 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 (A.2) 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.9) 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.9.

Table A.10: 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

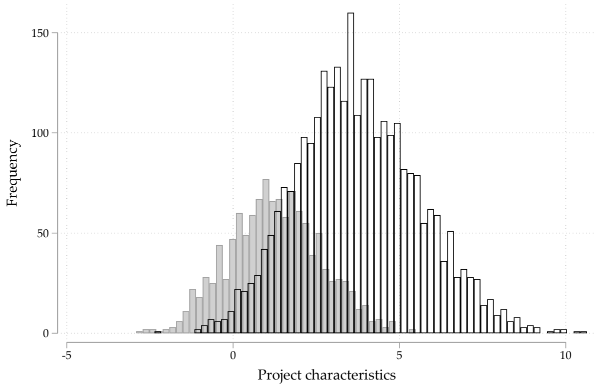
Note: see Table A.8. % reject $\beta \geqslant 1$ is the percentage of replications in which the null of full additionality is rejected at a 5% significance level.

The downward bias is still apparent in Table A.10 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 $\mu_c = 2$ 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).

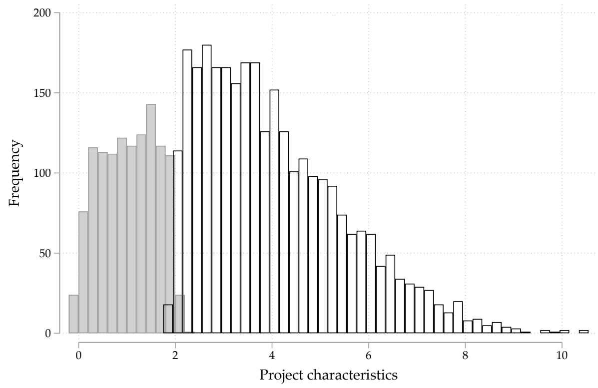
Our DGP creates a forgiving test-bed for difference and system GMM estimators, yet relatively persistent shifts over time in a country's average expected returns are enough for moment conditions to be violated and for these estimators to yield unreliable results. Adding other realistic features, like 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.

A.4 Firm-level results

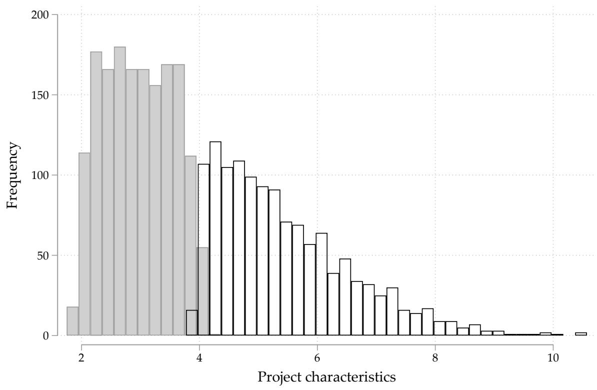
Figure A.1: Inferring additionality from firm-level data: additional examples



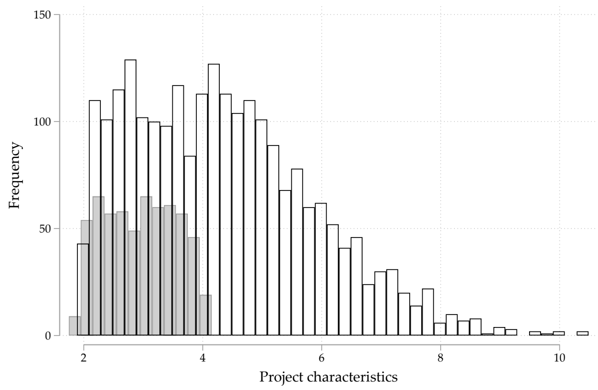
(a) Full additionality, large errors



(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 A.1a the variance in the non-observable part of expected returns is higher ($\sigma_e = 1.5$) than in Figure A.1b ($\sigma_e = 0.1$); $DB = 0.2 * nC * nI$ in both figures. In the bottom row there is zero additionality: $dfi_{lo} = 2, dfi_{hi} = 4$. In Figure A.1c the budget is larger ($DB = 0.3 * nC * nI$) than in Figure A.1d ($DB = 0.1 * nC * nI$); $\sigma_e = 0.1$ in both figures.

Figure A.1 gives four additional examples to further illustrate how the approach discussed in the main text can yield misleading results. Both plots in the top row are from a DGP with full additionality. Nonetheless, in Figure A.1a 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 A.1b, the full additionality of DFI investments reveals itself clearly. In the bottom row, the DGP has zero additionality. In Figure A.1c 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 A.1b. As in Figure A.1b, 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 A.1d), 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 A.1b and A.1c, as well as Figures A.1a and A.1d) clearly shows the problems associated with inferring additionality from firm-level data.

References

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

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

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

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

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

Bun, M. J. G. and Sarafidis, V. (2015). Dynamic panel data models. In Baltagi, B. H., editor, *The Oxford Handbook of Panel Data*, chapter 3, pages 76–110. Oxford University Press, Oxford, United Kingdom.

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

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

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

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

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

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

Windmeijer, F. (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.