

Paying for Poverty Alleviation: A Case Study

12 January 2018

Draft Working Paper

Richard Sedlmayr †

Integrated poverty alleviation programs that deliver a package of transfers and customized services to the poor over an extended period of time have repeatedly demonstrated noteworthy reductions in extreme poverty. But their multifaceted nature makes it difficult to project cost-effectiveness at large scale: because of large room for variability in the quality of service delivery, fidelity and impact may suffer as interventions expand beyond the original study settings. To fund in an evidence-based manner without making unwarranted assumptions nor imposing constraints on the complexity of programming, donors may want to abandon the presumption that the available body of evidence will allow them to adequately project the cost-effectiveness of a given program *ex ante*. To do so without abandoning the principles of evidence-based funding, they can take advantage of yet-unavailable evidence that will be generated over the course of a project's implementation. This paper presents and discusses the design of an outcome funded challenge to the development sector.

Keywords: aid effectiveness, external validity, evidence-based policy, results-based financing, delivery science

JEL Codes:

Author Affiliation: † Blavatnik School of Government and St Antony's College, University of Oxford.
richard.sedlmayr@sant.ox.ac.uk

Acknowledgments: This work is prepared in the context of an ongoing project involving multiple parties. My contributions included data analysis; donor advisory; participation in the design of the institutional setup; and the design of the payment function. The work could not have emerged without the thought partnership of Avnish Gungadurdoss, Douglas Emeott, Juan Camilo Villalobos, and Jan Witt of Instiglio, a nonprofit organization that promotes results-based financing in development and has the overall responsibility for the planning and management of the project. I also thank Dianne Calvi, Celeste Brubaker, and Caroline Bernadi at Village Enterprise; Jessica Cartwright and Radana Crhova at the UK Department for International Development; Duc Tran and Joaquin Carbonell at the US Agency for International Development; as well as Wesley Panek and Alice Gugelev at Global Development Incubator.

Transparency Statement: Supplementary materials, data, and code are accessible on the Open Science Framework (osf.io/mp5ns).

Why Outcome Payments? Some Background Aid Effectiveness and Evidence

Economic development grants are peculiar transactions: the benefits are not meant to accrue to the grantor (typically a donor) nor the grantee (typically a program implementer), but to third parties. In such settings, what forces will steer the allocation of resources? The intended beneficiaries, who are typically poor and disenfranchised, participate in the allocation process neither as market players nor as voters, so development practice relies on the transacting parties to represent their interests by proxy.

It is perfectly possible that these are truly concerned with their mandate: after all, altruism and complementary desires for reputation and self-respect are well-established drivers of human behavior that can dominate nonsocial incentives (Batson & Shaw, 1991; Fehr & Fischbacher, 2003; Ashraf, Bandiera, & Jack, 2014). But intentions alone do not adequately predict social outcomes. Even when agents are entirely self-oriented, their interactions in functioning markets can lead, as if by “invisible hand”, to socially desirable resource allocations. Conversely, prosocial preferences do not guarantee this if the right conditions are not in place. The manner in which the motivations of individuals interact and ultimately translate into social outcomes are governed by procedural rules, enshrined in institutions (North, 2000; Fukuyama, 2011).

On this front, aid has been subject to scathing criticism. It is routinely portrayed as flawed by design: unaccountable to the results it brings about, and effective only in undermining the local foundations of development (Bauer, 1972; Easterly, 2007; Moyo, 2009; Deaton, 2013). Such descriptions are frequently invoked in attempts to explain the failure of many empirical studies to corroborate evidence of aid effectiveness at the macro level (Doucouliagos & Paldam, 2009).

Criticisms of the institutional shortcomings of aid are widely acknowledged even by its proponents (Kenny, 2008, 2012). Diverse movements within development practice attempt to proactively influence its procedural rules. These efforts can be organized under guiding principles that prescribe (in a stylized manner, without necessarily claiming universality or finality) how charitable and aid grants should be allocated to expand their cost-effectiveness.

Two such principles are discussed here. One is the call that funding should be guided by knowledge generated through empirical science; the other holds that it should be tied to outcomes achieved. Both of these advocate for funding to be linked to evidence, generated by the means of impact evaluation. But the first defines this evidence in a prospective sense, used to project what will happen in the future; the other uses evidence retrospectively, to ascertain what happened in the past.

Use of Evidence for Projection

There is widespread support for the notion that development policy should be designed in the light of evidence that has been generated through rigorous empirical, especially experimental, research (Banerjee et al., 2006). The underlying idea is that administering micro-level interventions with sufficient granularity can allow for the identification of causal effects; that such evidence cumulates into a greater body of knowledge; and that this can in turn be used to adapt the development sector, accelerating effective approaches and dropping ineffective ones (Banerjee & Duflo, 2011). This calls for evidence to have a number of attributes:

Generalizability. A central challenge involves the external validity of available results. Even methods that go to great lengths to identify causality in a given setting (such as randomized evaluations, which subject the entire implementation process to scientific protocols) may not provide any elevated insight into the factors that moderated the impacts. Without such awareness, it is difficult to make informed extrapolations into new settings where moderators may differ (Cartwright & Hardie, 2012; Deaton, 2010). A theoretical grasp of the barriers to progress, and of the mechanisms through which interventions unfold impact, is necessary for the appropriate use of evidence in policymaking (Bates & Glennerster, 2017; Williams, 2017).

Actionability. More prosaic concerns pertain to the practicalities of translating research insights into action. Much research is poorly suited to the decision needs of practitioners: for instance, an imbalance between the efforts to establish impacts (often considered of academic interest) and those to establish costs (rarely considered of academic interest) may draw attention to detectable impacts as opposed to cost-effectiveness. More generally, the capacity of policymakers to digest academic literature may be limited, and absent an entry point into their minds, even insights that may be relevant to them in principle may pass them by in practice. Calls to interpret experimental insights only in the light of theoretical frameworks could raise the demands on such capacity.

Unbiasedness. Another approach to deepen generalizability involves the expansion of the body of complementary studies. However, even a voluminous body of research may fail to provide a sound basis for forming cost-effectiveness expectations, as diverse forces can skew its representativeness. For instance, there may be a bias in the selection of evaluation sites in favor of ones where capacities and conditions are more favorable than usual. Interventions that were unsuccessfully implemented may not yield useful evidence, introducing survivorship bias in the body of evidence. This is aggravated by publication bias driven by the norms and incentives of academia (Ioannidis, 2005). Research outlets may direct more attention towards results that are significant, intuitive, or entertaining, and less attention towards results that are insignificant or perplexing. In anticipation of such forces, researchers may choose to apply those analytical specifications that happen to yield welcome results, or simply deprioritize the

reporting of other results. There are indications that experimental results tend to be less biased by statistical reporting conventions than those of other empirical approaches, though they are not immune (Brodeur, Lé, Sangnier, & Zylberberg, 2016; Chan, Hrobjartsson, Haahr, Gotzsche, & Altman, 2004).

Use of Evidence for Retrospection

Another approach to aid effectiveness involves the creation of feedback loops that force a reckoning with programmatic achievements, or lack thereof. This may involve withholding funding from grantees until the desired results materialize; in other words, paying for success. To enable this, the grantor must commit to a results-based payment schedule *ex ante*. The justifications for this approach involve either incentivization or selection (Lazear, 2003).

If we suggest that results-based financing serves the purpose of creating incentives, we imply moral hazard problems: that the objectives of principals and agents differ, and that some relevant information (such as the amount of effort exerted by the agent) are not symmetric between the two. In this framework, results-based payments can be viewed as “performance pay” to align the agent’s incentives with the interests of the principal. (While it is most straightforward to think of principals as donors and agents as implementers, moral hazard may apply within donor institutions as well. For example, donor staff who are insufficiently concerned with the interests of the poor may fail to discontinue payments to a nonperforming project in order to avoid discomfort. This too might be avoided through results-based arrangements, where the burden to decide on the continuation of investments does not lie with donors.)

If we posit that most development practitioners are in fact motivated to serve the poor as best they can, principal-agent problems resembling adverse selection can remain. As an illustration, let us assume that various implementers have diverse beliefs sets about the proper approach to solving a problem. Among them, one bad apple is unconcerned with social impact and instead concerned with raising money to perpetuate its own existence and therefore willing to uncritically align with donor-held beliefs. If donors are naïve about this and fund only based on alignment of an applicant’s approach with their own beliefs, then the bad apple will be selected for funding. Meanwhile, if donors are aware of their predicament, they can rely on results-based financing as a signaling device to identify those implementers who intend to execute a course of action that they truly believe to be cost-effective.

Results-based financing may be useful even when there are no concerns about a mismatch between the intentions of donors and implementers at all. A feedback loop that select approaches based on their ability to achieve successful outcomes may give rise to a blind watchmaker: an evolutionary system that spawns functional designs without the need for profound awareness and deliberate creation. Such processes are at work not only in the biological, but also the socioeconomic domain (Dawkins, 1976; Lynch, 1996; Beinhocker, 2007). For illustration, assume that implementers share an intrinsic motivation to solve a

problem, but differ in their “quality”: and that both they and donors are aware of a variation in quality, but cannot appraise it at the outset. In this case, it may be appropriate for donors to take an agnostic stance and challenge implementers to demonstrate outcomes, leaving it to the natural selection effects of results-based financing to gradually advance the standing of higher quality implementers. This is most intuitive if we think of implementers as having some innate quality trait. To be sure, the capability of implementers will in fact be determined in part by factors that are malleable in principle, such as the accuracy of their beliefs and the appropriateness of their approach; and if these were updated efficiently in response to deepening collective experience, the adaptation of the development sector would function just fine without financial contingencies. But as discussed in the previous chapter, collective knowledge does not adapt to deepening experience as flawlessly as one might wish. Evolutionary forces may remedy this by leading the sector to adapt even to those lessons that, for whatever reason, are not efficiently codified into evidence, diffused, and learned. These might involve the kinds of nuanced, localized, and fleeting insights that emerge as implementers “muddle along” in trial-and-error (Andrews et al., 2013). In this reasoning, and all else unchanged: in domains where the room for fundamental knowledge has been narrowed and critical open knowledge gaps pertain to more complex and context-specific questions, results-based financing may be especially sensible.

Criticisms of results-based financing are diverse (Clist, 2016):

Motivational Effects. Performance pay can crowd out the intrinsic motivation by undermining the presumption of mutual trust that can otherwise govern the execution of incomplete contracts (Frey, 1997; Bénabou & Tirole, 2006; Falk & Kosfeld, 2006). Of additional concern, there is evidence to suggest that once a social contract is damaged through the introduction of an ill-conceived pecuniary incentive, it may not be restored later by the removal of the incentive (Gneezy & Rustichini, 2000).

Costs. Results-based financing can create costs not found in other arrangements. One of these involves the allocation of risk. Contingent payments reduce the predictability of cash flows for implementers, who may be risk averse (Miller & Babiarz, 2014). If we interpret this risk as being shifted away from less risk-averse donors, one could interpret a resulting net risk premium as a cost to the project.¹ Results-based

¹ Such argumentation is complicated by two factors. On the one hand, the concern may be mitigated by the fact that there are markets for risk: in the context of so-called social impact bonds, commercial or socially minded investors provide risk capital to the implementer, thereby shrinking or eliminating the net risk premium. On the other hand, the presumed risk reduction for donors from results-based financing is easily overestimated: for ones who are mandated to make predictable payouts – e.g., private foundations who, under US tax provisions, are penalized for not meeting minimum payout requirements – a reduction in the predictability of cash outflows can be unwelcome. So while results-based financing may shift performance risk from donors to implementers, cash flow related risk will tend to increase for both.

financing may also increase transaction costs (Eldridge & Tekolste, 2016). By removing a presumption of strategic alignment, the transacting parties may be more inclined to invest in legal and communication expenses to prepare for or engage in disagreements. The introduction of additional partners (such as investors, technical service providers, or co-funders) can further increase transaction costs, especially when all parties contract jointly (Gong, Shenkar, Luo, & Nyaw, 2007). At the time of writing, even after the initiation of dozens of social impact bonds over the last six years amounting to hundreds of millions of dollars in transaction volume, there has yet to emerge one such deal that is created without a subsidy to de-risk investments or cover transaction costs (Floyd, 2017).

Gaming Concerns. Once a measure is defined as a target, incentive processes can corrupt it, as intuitively illustrated through the story that a bounty on rat tails that was meant to control rat populations in Hanoi instead encouraged citizens to breed them (Campbell, 1979; Goodhart, 1975; Vann, 2003). Short of entirely backfiring, incentives can distort a multitasking implementer's priorities towards those tasks whose success is more easily measured, away from less tangible ones (Holmstrom & Milgrom, 1991). The gaming of performance incentives is empirically well established in public policy contexts (Courty & Marschke, 2008; Dixit, 2002; Jacob & Levitt, 2003).

One appropriate setting for results-based approaches may involve the scaling of poverty alleviation programs.

Why Outcome Payments in Poverty Alleviation? On the Policy Challenge of Achieving Scale

So-called “ultra-poor graduation” programs target the poorest in low-income countries and offer them an integrated program including transfers, training, and mentorship in an attempt to help them establish sustainable microenterprises. A meta-study by Banerjee et al. (2015), which summarizes six randomized pilot studies of such interventions, is widely cited as a milestone in development research, indicating that sustained poverty relief is achievable through a micro-level intervention across contexts. Let us review the value of this work for policymaking in the light of the aforementioned difficulties in using evidence for projection.

Actionability

The insights are regarded as highly actionable, e.g., central to the rationale behind the 2017 establishment of a new unit and associated multi-donor trust fund at the World Bank with the mission to promote graduation-style policy approaches in social protection and economic development projects worldwide (Consultative Group to Assist the Poor, 2017).

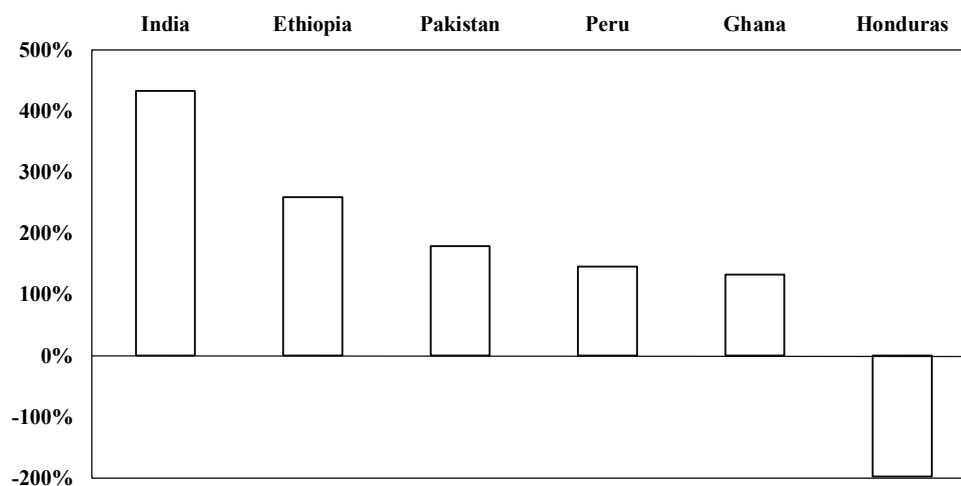
Unbiasedness

The paper makes a set of projections about the sustainability of economic effects that could be seen as aggressive assumptions, but have started since started to gain empirical support (see e.g. Banerjee, Duflo, Chattopadhyay, & Shapiro, 2016). It excludes some relevant randomized trials of graduation programs; one of these was large and successful (Bandiera et al., 2016), another was smaller and did not yield sustained impact (Bauchet, Morduch, & Ravi, 2015). There also exist randomized evaluations that are similar in spirit to graduation programs, but are not labeled as such and not included in the review (see e.g. Blattman et al., 2016). Overall, the rates of cost-effectiveness in Banerjee et al. (2015) appear representative of the available body of evidence. As this meta-study only includes studies that were fully implemented, a projection of expected cost-effectiveness of similar programs may need to account for the possibility of implementation failure.

Generalizability

The central difficulty in forming expectations about the cost-effectiveness of future graduation-type programs emanates from the substantial heterogeneity in the results: the program was dramatically more cost-effective in some settings than others. Figure 1 presents cost-benefit metrics using the paper’s primary set of assumptions.

Figure 1: Cost-Benefit Ratios Found in Six “Graduation” Sites



Note: For underlying assumptions, consult the data source: Banerjee et al. (2015).

The authors discuss this heterogeneity. In the least successful site, the transferred assets depreciated at high rates, which the authors interpret as an indication that the quality of the program’s asset component may be an important success factor. But they do not claim that this is the only reason for the observed heterogeneity – with good reason: there may be numerous other moderators. While some of these will

remain unknown, others are obvious. For instance, while we cannot establish to what degree the quality of non-tangible asset component (i.e., the intangible human capital created through training and coaching) differed across sites, we know that to the extent these activities have any merit at all, their impacts must be moderated by the quality of implementation: training and coaching will not work well if they are executed poorly. This is a serious concern if the quality of implementation is negatively correlated with the scale of implementation – a conjecture with some empirical support (Bold, Kimenyi, Mwabu, Ng'ang'a, & Sandefur, 2013).

Summing up: available evidence demonstrates that it is possible to cost-effectively reduce extreme poverty across contexts and in the long run. It does not however provide high confidence that a specific approach will again achieve this in a given setting or at very large scale. The core motivation of the work presented here is to anticipate this and explore a possible remedy, involving the retrospective use of yet-unavailable evidence.

The Case of The Village Enterprise Outcomes Fund

The challenge is addressed in the context of a micro-enterprise development intervention administered by an organization called Village Enterprise, a small non-profit headquartered in the United States. It has an annual operating expense budget between USD 1.2M (2013) and 2.1M (2016), which it spends on microenterprise programming in rural regions of Uganda and Kenya. The program is similar to so-called ultra-poor graduation in that it targets the poorest (ordinarily between the poorest quartile and the poorest half of a given village, as measured through a participatory wealth ranking and validated through a Progress-out-of-Poverty survey) and offers them an integrated package of transfers, training, and mentorship. The program involves the formation of both business groups (target size: three households) and savings groups (target size: 30 households). Transfers are provided in the form of cash injections to business groups and account to approximately 1/3 of direct program costs; a larger portion of costs is associated with training and mentoring. The theory of change holds that the implementer *provides households with a tangible asset (via a transfer), as well as an intangible asset (i.e., human capital, generated via training and mentorship) that is meant to help the household maintain and derive value from the tangible asset. This leads households to engage in more productive economic activities, generate higher income, and achieve sustained improvements in living standards.*

Between 2013 and 2016, Village Enterprise was involved in a randomized trial evaluating the impacts of diverse components and variants of its program. Results suggest that Village Enterprise succeeded in reducing poverty at rates of cost-effectiveness that compare favorably to the rates found in diverse evaluations of ultra-poor graduation programs. They also provide some indication that intangible

extensions can alleviate poverty more cost-effectively than if they were monetized and provided in the shape of plain transfers.

At the time of writing, a group of donors has agreed to fund an expansion Village Enterprise's programming. Rather than projecting with confidence that specific rates of cost-effectiveness will be achieved in the future, donors have committed funds to a results-based challenge, offering to compensate Village Enterprise as a function of outcomes that will be measured over the course of implementation. Specifically, the challenge is framed as follows:

As the body of available evidence on integrated poverty alleviation programs deepens, priorities shift to delivery science. What rules and procedures should be followed to allow integrated poverty alleviation programs to be delivered effectively and at large scale? This is less a question of generalizable insight (say, whether a class of programs is able to overcome fundamental constraints to prosperity across contexts) but of practical localized decision-making (say, on how to cost-effectively recruit, train, motivate, and deploy qualified trainers and coaches in specific regions of Uganda). Donors recognize their own limitations in grasping the success factors on such issues and rely on results based financing to shift expanded responsibility to entrepreneurially minded implementers.

Numerous project details are explained in a forthcoming design memo by Instiglio (2018); the following chapters focus on the specific approaches taken to address the aforementioned concerns about results-based financing in its context.

Dealing with Motivational Concerns

While results-based financing may adversely affect the implementer's intrinsic motivation in some settings, this may not be a necessary consequence. Consider the mechanism through which motivation crowding is thought to work. It takes place when a principal's actions are perceived as controlling (Frey, 2017) and cast doubt on the nonpecuniary motives or trustworthiness of the agent (Falk & Kosfeld, 2006). It is this control, and the presumed motivation behind it, that is thought to lead to a loss of intrinsic motivation: for this reason, the adverse performance effects of motivation crowding are in fact referred to as "hidden costs of control".

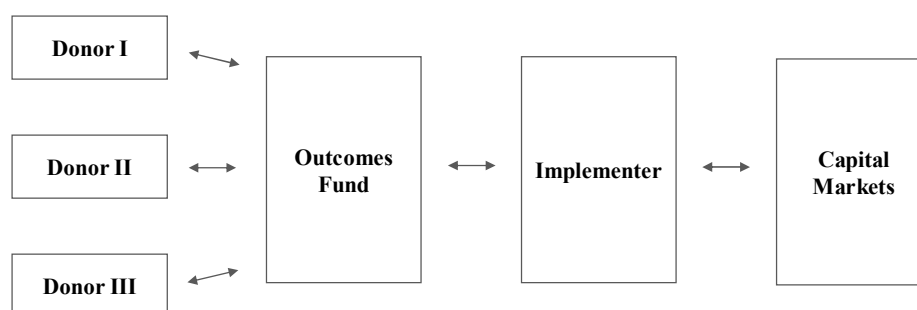
If this is true, the costs might be avoided through thoughtful communication that deliberately abandons a presumption of moral hazard, or even the mental model of principal-agent interactions more generally. This is a central motivation of how the challenge is presented (see above). It does not frame implementers

as morally dubious subordinates to be controlled, but as pioneering entrepreneurs to be empowered. As such, it is unlikely to crowd out intrinsic motivation.²

Dealing with Cost Concerns

The contractual framework depicted in Figure 2 is aimed at mitigating costs associated with inefficient risk allocation, as well as the transaction costs associated with complex joint ventures (Gong et al., 2007). All contracts are bilateral: no contract is signed by more than two parties.

Figure 2: Contractual Framework



Note: Arrows depict contractual relationships.

Donors pay into a single outcomes fund, which in turn contracts with the implementer. By paying into this fund at the outset of the project, donors can signal a commitment that may mitigate potential perceptions of counterparty risk, possibly lowering capital costs for the implementer. The existence of the outcomes fund may also reduce transaction cost: the contract between the outcome payer and outcome payee (i.e., implementer) is the single most sensitive and complex transaction in the arrangement, and the pooling of donors allows for it to be executed once as opposed to three times.

The donor commitments precede the implementer's campaign to raise working capital and potentially mitigate financial risk exposure by raising investment funds from the capital markets. The potential involvement of investors gives the project the central feature of so-called social (or development) impact bonds, though the financial dealings to raise working capital are left to the implementer. As "customers of

² This speaks to a more fundamental scientific debate. It has been proposed that institutional agents (such as Village Enterprise) are generally be less susceptible to experiencing motivation crowding from results-based financing than individual agents are, but both theoretical and empirical corroborations are weak (Clist, 2016). Perhaps conditional payments are perceived as more typical of institutions while trust-based transactions are perceived as more typical of individuals; if so, conditional payments may be less readily interpreted in an insulting and disempowering manner by institutional agents.

impact”, outcome payers are not at all involved in the financing of the production process. This mirrors a results-based financing tool called the advance market commitment (Berndt et al., 2006), and most private sector transactions.

Dealing with Gaming Concerns

One way in which the implementer could corrupt the process might involve the preparation of respondents to respond to the independent evaluation survey in some way. This can be prevented through preparations to detect and penalize such behavior.

A more difficult question pertains to the design targets in a manner that does not incentivize the provider to achieve something else than the ultimately intended objectives. For this purpose, let us review the payment formula. It emanates the program’s theory of change, as described above.

In the interest of reducing working capital requirements and thereby capital costs, the parties strive to agree on payments as soon as justifiable. At the time of programming (period zero), the impact of the asset transfer on the net asset position of beneficiaries can already be known. (It can be established without the need for a household survey: so long as an independent validation exercise can ascertain that the transfer occurred, we can be satisfied that its impact on the net asset position equals its nominal amount.) There is therefore no justification for withholding a payment in period zero:

$$(I) \quad P_0 = a_0$$

Here, P_0 is the payment that is made by the outcomes fund to the implementer and a_0 is the impact of the transfer payment on the net asset position of beneficiaries, which equals the nominal asset transfer.

To justify additional payments, we want to establish if the implementer has generated value beyond the cost of the tangible transfer. This is not done in period zero, as the human capital portion is not easily valued as a stock. (One creative approach that could achieve this in principle is to establish respondents’ willingness to pay for human capital; see Shapiro, 2017. It is however questionable if such assessments will be anywhere near accurate; see Jensen, 2010.)

We wait until period m to establish if the implementer’s activities generated additional benefits and value these benefits in the form of periodic flows. We choose the flow of periodic incremental consumption as the target outcome: it is less noisy than the alternative measure income—partly because it reflects intertemporal smoothing and partly because it is easier to measure accurately (Deaton, 1997; Meyer & Sullivan, 2003). P_m is the payment made in period m (i.e., the period of outcome measurement):

$$(II) \quad P_m = -a_0(1+r)^m + \sum_{i=1}^{\infty} c_i \left(\frac{1}{1+r}\right)^{i-m}$$

Here, r is a discount rate (compounded periodically) and c_i is the incremental consumption in period i . We cannot afford to measure consumption indefinitely, so P_m needs to be adjusted to rely only on measurables. For this purpose, we estimate c_m (i.e., consumption in period m), as well as a rate at which consumption effects change over time. For finite values, the rate of consumption growth cannot be higher than the discount rate. We assume that it is negative and define a rate of decline d , yielding:

$$(III) \quad P_m = -a_0(1+r)^m + \sum_{i=1}^{\infty} c_m \left(\frac{1-d}{1+r}\right)^{i-m}$$

For intuition and later simplification, we distinguish short-term values (in periods 1 through m) and long-term values (beyond period m):

$$(IV) \quad P_m = -a_0(1+r_S)^m + c_m \sum_{i=1}^m \left(\frac{1-d_S}{1+r_S}\right)^{i-m} + c_m \sum_{j=m+1}^{\infty} \left(\frac{1-d_S}{1+r_S}\right)^{j-m}$$

The first term is the value of the asset, which was already compensated at time zero and is subtracted in period m ; the second term is the value of accumulated consumption between periods zero and m ; and the third term is a perpetuity that captures the value of consumption beyond period m . Short- and long term rates of discount and decline are denoted through sub-scripts S and L.

The payment formula has important merits. It achieves the central objective of aligning the incentives of the implementer with permanent consumption stream of the beneficiary. To see this, combine (1) and (II) into a present value of payments in period zero:

$$(V) \quad NPV_0(Payments) = P_0 + \frac{P_m}{(1+r)^m} = \sum_{i=1}^{\infty} \frac{c_i}{(1+r)^i}$$

In defining present value of consumption as the central objective, the payment function identifies a metric that is widely applicable for micro-level poverty alleviation programming. (Only projects that have explicitly different objectives than sustained poverty reduction—say, non-economic or humanitarian ones—can plausibly claim exemption from this framework). To see why this is important, take the ongoing debate about the merits of unconditional cash transfers as a tool for development. Advocates argue that such transfers expand the agency of beneficiaries in the allocation of development resources, allowing them to exercise choice. They also point to evidence that when provided with such freedom, the poor do not act in a manner that can be characterized as wasteful or irresponsible (Blattman & Niehaus, 2014). A counter-argument is that many other failures than ones related to capital markets have been established as constraints to economic progress, so there are reasons to believe that it is possible to “do better than cash”. The challenge is ultimately an empirical one: choosing between unconditional transfers and alternative interventions calls for evidence on relative impact. But different development programs

can yield diverse possible outcomes and different patterns of outcome realization over time; to benchmark approaches against each other (or even against their own cost), we must define some utility function through which outcomes and their time patterns can be valued. The payment function achieves this.

Simultaneously, it provides the implementer with justifiable working capital as early as period zero, when a lower bound impact can already be established and compensated. Payment P_0 is by definition lower than the cost of the program; any strategy to derive value exclusively from this payment must therefore, in expectation, be unprofitable. Payment P_m claws back the initial payment before any new information emerges and redefines the implementer's target as the present value of incremental consumption.

Some caveats must be pointed out. For example, reasonable actions taken to operationalize this framework can distort incentives, upsetting equation (V). In the case of the Village Enterprise project, the parties opted to simplify the payment formula for enhanced intuition. Both the short-term discount rate r_s and the short-term rate of consumption decline d_s are defined as zero. (The latter implies that the incremental consumption measured in period m is representative of incremental consumption to date.) This allows us to simplify (IV) to:

$$(VI) \quad P_m = -a_0 + c_m m + c_m x$$

The first two terms are no longer discounted, which defines capital as free for the duration of the project. Resulting distortions are likely unimportant in an environment of low capital costs. But more serious concerns relate to the third term. Perpetuity x can be simplified to multiplier $x = (1 - d_L)/(r_L + d_L)$ without distortion, but its constituent components cannot be left to vary freely. If d_L is very high, this can create a liability for the implementer; if it is very low or negative, it can raise the payments beyond the capacities of the outcome payers. For $d_L \leq -r_L$, multiplier x must remain undefined. Some bounds are therefore defined for allowable values of d_L . Further bounds are defined on allowable payments. (See Appendix 1 for details on bounding rules.) Such bounds can alter the risk-adjusted calculus of the implementer. For intuition, assume that a self-oriented implementer chose to maximize risk-adjusted returns in the light of the agreement that P_m cannot be negative. In this case, a strategy to provide a large tangible asset and a no intangible asset would have little downside for the implementer. As the tangible portion is immediately compensated through P_0 , the maximum loss would amount to administrative and delivery costs associated with the transfer, which may be low. Meanwhile, payment P_m could be positive out of pure chance, and risk is substantial (see Appendix 2 for simulations). The desire to achieve a high signal-to-noise ratio explains why the future consumption and asset estimates will again be derived from a large-scale randomized trial. All else unchanged: the lower the variance of expected effects, and the less constraining any bounds, the cleaner the link between outcomes and payment, and the lower the room for

distortion. In the specific case of the Village Enterprise project, the selected bounds are thought to have net benefits: they are thought to cap important tail risks to both donor and implementer, without distorting actions far from what the implementer considers optimal (partly because the project's potential demonstration value puts some reputational capital at stake).

This points to an additional caveat. One potential appeal of results-based financing not discussed above is the vision that different results-based transactions could ultimately be synthesized into ever more competitively priced "outcome menus" for donors to consult in their choices. However, if implementers subsidized their work through other means than the results based payments – say, because of reputational concerns, or some other benefit that is associated with success but not captured by the results based payments – then the transaction prices may underestimate the true investment in (and therefore social cost of) the outcomes. This may be hard to overcome, as the appropriate costing of managerial and indirect subsidies is a tall task. As it is not even attempted in the Village Enterprise project, the cost-effectiveness to the project from a social perspective will not be quantified with precision.

Conclusion

Donors can engage in results-based financing for other reasons than to incentivize implementers to perform. For instance, they may lack an adequate basis for confident predictions about the expected impacts specific activities and restrict their activities to creating demand for proven outcomes, or "ends". This shifts the burden of efficiently selecting and adapting "means" to the supply of the market for development outcomes. This is in turn efficient if the body of implementers on the "supply side" is better able to adapt to emerging insights than donors are on the "demand side", which may be true for a number of reasons. First, implementers may generate context-specific insights that are simultaneously important for success and important to freely adapt to, yet difficult to codify into a more broadly useful body of knowledge. Second, even if individual implementers failed to deliberately adapt for themselves, the evolutionary forces of results-based financing would select for implementers whose approaches happen to be more appropriate.

Results-based financing comes with new difficulties, e.g., concerns about (a) costs, (b) motivational effects, and (c) possible gaming. We discuss a project's strategy for addressing each of these concerns, specifically (a) the establishment of an outcomes fund and associated contracting structure; (b) behavioral design building on insights from research on motivation crowding; and (c) a payment function that aligns targets with ultimately desired outcomes while remaining sensitive to capital and measurement costs. These strategies provide a template that is, in principle, widely applicable for micro-level poverty alleviation funding strategies.

Bibliography

- Andrews, M., Pritchett, L., & Woolcock, M. (2013). Escaping Capability Traps Through Problem Driven Iterative Adaptation (PDIA). *World Development*, 51, 234–244.
<https://doi.org/10.1016/j.worlddev.2013.05.011>
- Ashraf, N., Bandiera, O., & Jack, B. K. (2014). No margin, no mission? A field experiment on incentives for public service delivery. *Journal of Public Economics*, 120, 1–17.
<https://doi.org/10.1016/j.jpubeco.2014.06.014>
- Bandiera, O., Burgess, R., Das, N., Selim, G., Rasul, I., & Sulaiman, M. (2016). Labor Markets and Poverty in Village Economies. *Mimeo*, (February).
- Banerjee, A., Deaton, A., Lustig, N., Rogoff, K., & Hsu, E. (2006). *An Evaluation of World Bank Research, 1998 - 2005*. Washington DC.
- Banerjee, A., & Duflo, E. (2011). *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty*. PublicAffairs. Retrieved from <https://books.google.com/books?id=Tj0TF0IHlyAC>
- Banerjee, A., Duflo, E., Chattopadhyay, R., & Shapiro, J. (2016). The Long term Impacts of a “Graduation” Program: Evidence from West Bengal. *Working Paper*, (September).
- Banerjee, A., Duflo, E., Goldberg, N., Karlan, D., Osei, R., Pariente, W., ... Udry, C. (2015). A multifaceted program causes lasting progress for the very poor: Evidence from six countries. *Science*, 348(6236). <https://doi.org/10.1126/science.1260799>
- Bates, M. A., & Glennerster, R. (2017). The Generalizability Puzzle. *Stanford Social Innovation Review*, (Summer). Retrieved from https://ssir.org/articles/entry/the_generalizability_puzzle
- Batson, C. D., & Shaw, L. L. (1991). Evidence for altruism: Toward a pluralism of prosocial motives, 2(2), 107–122. https://doi.org/10.1207/s15327965pli0202_1
- Bauchet, J., Morduch, J., & Ravi, S. (2015). Failure vs. displacement: Why an innovative anti-poverty program showed no net impact in South India. *Journal of Development Economics*, 116, 1–16.
<https://doi.org/10.1016/j.jdeveco.2015.03.005>
- Bauer, P. (1972). *Dissent on Development*. Cambridge, MA: Harvard University Press.
- Beinhocker, E. D. (2007). *The Origin of Wealth*. Cambridge, MA: Harvard Business Review Press.
- Bénabou, R., & Tirole, J. (2006). Incentives and Prosocial Behavior. *American Economic Review*, 96(5), 1652–1678. <https://doi.org/10.1017/CBO9781107415324.004>

- Berndt, E. R., Glennerster, R., Kremer, M. R., Lee, J., Weizsacker, G., Williams, H., & Levine, R. (2006). *Advance Market Commitments for Vaccines Against Neglected Diseases : Estimating Costs and Effectiveness*.
- Blattman, C., Green, E. P., Jamison, J., Christian, M., Annan, J., Carlson, N., ... Peck, R. (2016). The Returns to Microenterprise Support among the Ultrapoor: A Field Experiment in Postwar Uganda. *American Economic Journal: Applied Economics*, 8(2), 35–64.
<https://doi.org/http://dx.doi.org/10.1257/app.20150023>
- Blattman, C., & Niehaus, P. (2014). Show them the money. *Foreign Affairs*.
- Bold, T., Kimenyi, M., Mwabu, G., Ng'ang'a, A., & Sandefur, J. (2013). Scaling-up What Works: Experimental Evidence on External Validity in Kenyan. *Centre for Study of African Economies Conference*, 44, 1–47.
- Brodeur, A., Lé, M., Sangnier, M., & Zylberberg, Y. (2016). Star wars: The empirics strike back. *American Economic Journal: Applied Economics*, 8(1), 1–32. <https://doi.org/10.1257/app.20150044>
- Campbell, D. T. (1979). Assessing the impact of planned social change. *Evaluation and Program Planning*, 2(1), 67–90. [https://doi.org/10.1016/0149-7189\(79\)90048-X](https://doi.org/10.1016/0149-7189(79)90048-X)
- Cartwright, N., & Hardie, J. (2012). *Evidence-Based Policy: A Practical Guide to Doing It Better*. Oxford University Press. Retrieved from <https://books.google.com/books?id=SwU8LgEACAAJ>
- Chan, A. W., Hrobjartsson, A., Haahr, M. T., Gotzsche, P. C., & Altman, D. G. (2004). Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA : The Journal of the American Medical Association*, 291(20), 2457–2465. <https://doi.org/10.1001/jama.291.20.2457> [pii]
- Clist, P. (2016). Payment by results in development aid: All that glitters is not gold. *World Bank Research Observer*, 31(2), 290–313. <https://doi.org/10.1093/wbro/lkw005>
- Consultative Group to Assist the Poor. (2017). CGAP Announces New Platform for Graduation Approach. Retrieved November 11, 2017, from <http://www.cgap.org/news/cgap-announces-new-platform-graduation-approach>
- Courty, P., & Marschke, G. (2008). A general test for distortions in performance measures. *Review of Economics and Statistics*, 90(3), 428–441. <https://doi.org/10.1162/rest.90.3.428>
- Dawkins, R. (1976). *The Selfish Gene*. Oxford: Oxford University Press.

- Deaton, A. (1997). *The analysis of household surveys: A microeconomic approach to development policy*. Washington DC: World Bank Publishing Group. <https://doi.org/doi:10.1596/0-8018-5254-4>
- Deaton, A. (2010). Instruments, Randomization, and Learning about Development. *Journal of Economic Literature*, 48(2), 424–455. <https://doi.org/10.1257/jel.48.2.424>
- Deaton, A. (2013). *The Great Escape: Health, Wealth, and the Origins of Inequality*. Princeton: Princeton University Press.
- Dixit, A. (2002). Incentives and Organizations in the Public Sector An Interpretative Review. *The Journal of Human Resources*, 3743(18), 696–727. <https://doi.org/10.2307/3069614>
- Doucouliaagos, H., & Paldam, M. (2009). The aid effectiveness literature: The sad results of 40 years of research. *Journal of Economic Surveys*, 23(3), 433–461. <https://doi.org/10.1111/j.1467-6419.2008.00568.x>
- Easterly, W. (2007). *The White Man's Burden: Why the West's Efforts to Aid the Rest Have Done So Much Ill and So Little Good*. Penguin Group.
- Eldridge, M., & Tekolste, R. (2016). *Results-Based Financing Approaches: Observations for Pay for Success from International Experiences*.
- Falk, A., & Kosfeld, M. (2006). The Hidden Costs of Control. *The American Economic Review*, 96(5), 1611–1630. <https://doi.org/10.2307/30034987>
- Fehr, E., & Fischbacher, U. (2003). The nature of human altruism. *Nature*, 425(6960), 785–791. <https://doi.org/10.1038/nature02043>
- Floyd, D. (2017). Briefing on Social Impact Bonds Social Impact Bonds An Overview of the Global Market for Commissioners and Policymakers, (April). Retrieved from http://socialspider.com/wp-content/uploads/2017/04/SS_SocialImpactReport_4.0.pdf
- Frey, B. S. (1997). Not Just for the Money: an Economic Theory of Personal Motivation. *Wiley Online Library*, 80, 77–78.
- Frey, B. S. (2017). Policy consequences of pay-for-performance and crowding-out. *Journal of Behavioral Economics for Policy*, 1(1), 55–59.
- Fukuyama, F. (2011). *The Origins of Political Order*. New York: Farar, Straus and Giroux.
- Gneezy, U., & Rustichini, A. (2000). A Fine is a Price. *The Journal of Legal Studies*, 29(1), 1–17. <https://doi.org/10.1086/468061>

- Gong, Y., Shenkar, O., Luo, Y., & Nyaw, M.-K. (2007). Do multiple parents help or hinder international joint venture performance? The mediating roles of contract completeness and partner cooperation. *Strategic Management Journal*, 28(10), 1021–1034. <https://doi.org/10.1002/smj.626>
- Goodhart, C. (1975). *Problems of Monetary Management: The U.K. Experience* (Papers in Monetary Economics (Reserve Bank of Australia)).
- Holmstrom, B., & Milgrom, P. (1991). Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design. *Journal of Law, Economics, and Organization*, 7(special), 24–52. https://doi.org/10.1093/jleo/7.special_issue.24
- Instiglio. (2018). *Design Memo: Village Enterprise Development Impact Bond*. Bogota.
- Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, 2(8), 0696–0701. <https://doi.org/10.1371/journal.pmed.0020124>
- Jacob, B. A., & Levitt, S. D. (2003). Rotten apples: An investigation of the prevalence and predictors of teacher cheating. *The Quarterly Journal of Economics*, 118(3), 843–877. <https://doi.org/10.1162/00335530360698441>
- Jensen, R. (2010). The (Perceived) Returns to Education and the Demand for Schooling. *Quarterly Journal of Economics*, 125(2), 515–548. <https://doi.org/10.1162/qjec.2010.125.2.515>
- Kenny, C. (2008). What is effective aid? How would donors allocate it? *European Journal of Development Research*, 20(2), 330–346. <https://doi.org/10.1080/09578810802078704>
- Kenny, C. (2012). *Getting Better: Why Global Development Is Succeeding--And How We Can Improve the World Even More* (2nd Editio). New York: Basic Books.
- Lazear, E. P. (2003). Teacher incentives. *Swedish Economic Policy Review*, 10, 179–214.
- Lynch, A. (1996). *Thought Contagion: How Belief Spreads Through Society: The New Science Of Memes*. The Kluwer international series in engineering & computer science.
- Meyer, B. D., & Sullivan, J. X. (2003). *Measuring the Well-Being of the Poor Using Income and Consumption* (NBER Working Paper Series No. 9760). Retrieved from <http://www.nber.org/papers/w9760>
- Miller, G., & Babiartz, K. S. (2014). Pay-for-Performance Incentives in Low- and Middle-Income Country Health Programs. *Encyclopedia of Health Economics*, 457–466. <https://doi.org/10.1016/B978-0-12-375678-7.00126-7>

- Moyo, D. (2009). *Dead Aid*. New York: Farrar, Straus and Giroux.
- North, D. C. (2000). *Institutions, Institutional Change and Economic Performance*. New York: Cambridge University Press.
- Shapiro, J. (2017). *Benchmarking Development Programs: a Preference-Based Approach* (mimeo).
- Vann, M. G. (2003). Of Rats, Rice, and Race: The Great Hanoi Rat Massacre, an Episode in French Colonial History. *French Colonial History*, 4(1), 191–203. <https://doi.org/10.1353/fch.2003.0027>
- Williams, M. J. (2017). *External validity and policy adaptation: From Impact Evaluation to Policy Design* (BSG Working Paper Series No. BSG-WP-2017/019). Retrieved from <http://www.martinjwilliams.com/research>

Appendix 1: Bounds in the Village Enterprise Project

Two classes of bounds are introduced. One bound limits allowable measures of asset effects and associated consumption projections, so as to avoid undefined or extreme values. The other bound put caps and floors on allowable payments directly.

Bound 1: Scenarios of Growth / Decline

Scenario A: Asset effects have disappeared. Initially, assume pessimistically that by period m , tangible assets revert to the original levels, i.e., that no effect on tangible assets persists. $a_m \leq 0$ in this scenario. This implies that the asset that is critical to the theory of change has been consumed, and therefore does not warrant the expectation that a further stream of benefits will continue to accrue beyond period m . However, it is possible that up to the current date, the flow of benefits has already exceeded of the value of the asset, and at minimum, P_m should account for these benefits; the first two terms of (VI) therefore remain intact. In summary,

$$(VII) \quad x_A = 0$$

Scenario B: Asset effects have not fallen. Hopefully, the implementer can demonstrate that it has achieved a degree of sustainability that warrants a payment in excess of P_{m_A} . As the theory of change holds that tangible assets are necessary enablers of sustained change, the assumption of sustainability can only be warranted if $a_m \geq 0$. Ideally, if by month m assets have not fallen at all, i.e., $a_m \geq a_0$. If so, it is reasonable to expect that benefits will persist. We assume $d_{L_B} = 0$, implying:

$$(VIII) \quad x_B = \sum_{j=m+1}^{\infty} \left(\frac{1}{1+r_L}\right)^{j-m} = \sum_{j=1}^{\infty} \left(\frac{1}{1+r_L}\right)^j = \frac{1}{r_L}$$

Scenario C: Asset effects have diminished. In the expected scenario, asset effects by period m have neither fully persisted nor fallen to zero, but are lower than the initial asset transfer has been: $a_0 > a_m > 0$. This suggests that consumption effects are also declining in the future. To operationalize it for the context at hand, we must account for the fact that we must estimate rates of decline using one single survey. We can take advantage of the fact that the program's impact on assets at time zero is already known. Specifically, we assume that beyond period m , the rate of consumption decline corresponds to the observed rate of asset decline:

$$(IX) \quad d_{L_C} = 1 - \left(\frac{a_m}{a_0}\right)^{\frac{1}{m}}$$

This assumption implies constant returns to capital – implausible across the production function, but reasonable at the margin. Consequently:

$$(X) \quad x_C = \sum_{j=m+1}^{\infty} \left(\frac{1-d_L}{1+r_L} \right)^{j-m} = \frac{1-d_L}{r_L+d_L} = \frac{\left(\frac{a_m}{a_0} \right)^{\frac{1}{m}}}{r_L+1-\left(\frac{a_m}{a_0} \right)^{\frac{1}{m}}}$$

Bound 2: Payment Caps / Floors

P_m is further be restricted between two values to avoid tail risks arising to both donor and implementer. As a floor, P_m shall not be so low that it assigns excessive risk to the implementer; as a cap, P_m shall not exceed available outcome payments $P_{m_{cap}}$. In summary, the payment rule to be executed in period m is as follows:

$$\begin{aligned} & \text{IF } a_m \leq 0 \quad \rightarrow \text{ THEN } P_m = \text{Median} \left[P_{m_{floor}}, P_m(x_A), P_{m_{cap}} \right] \\ (XI) \quad & \text{IF } a_m \geq a_0 \quad \rightarrow \text{ THEN } P_m = \text{Median} \left[P_{m_{floor}}, P_m(x_B), P_{m_{cap}} \right] \\ & \text{IF } a_0 > a_m > 0 \quad \rightarrow \text{ THEN } P_m = \text{Median} \left[P_{m_{floor}}, P_m(x_C), P_{m_{cap}} \right] \end{aligned}$$

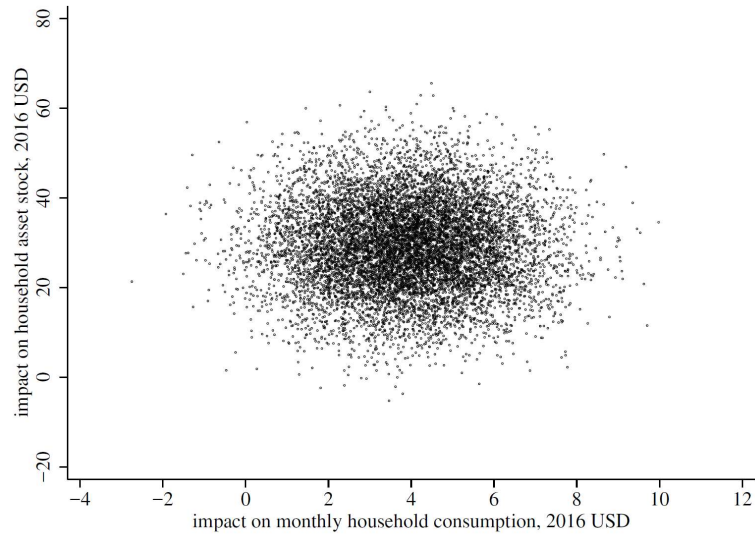
In the Village Enterprise project, the minimum and maximum payments per households in period m are:

$$P_{m_{floor}} = \text{USD } 0$$

$$P_{m_{cap}} = \text{USD } 265$$

Appendix 2: Simulations

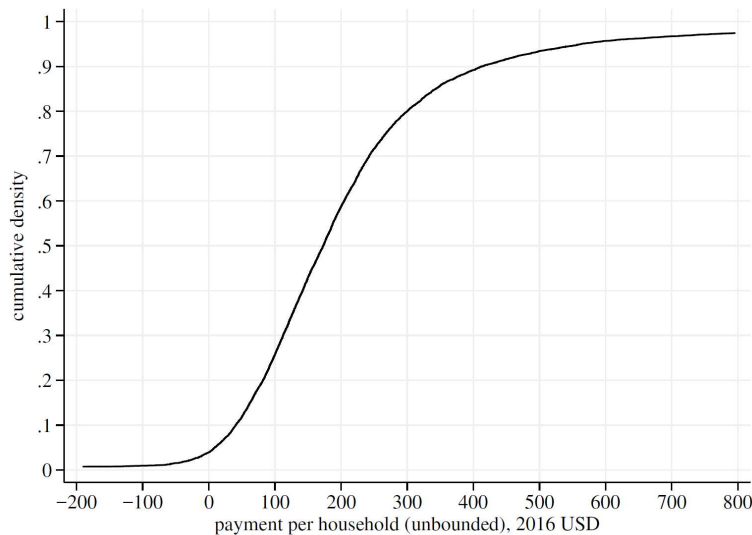
Figure 3: Simulated Effects using Existing Village Enterprise RCT



Explanation: This uses estimates and associated errors from the existing Village Enterprise RCT to simulate a distribution of “true” effects ($n=5,000$). Note that this simulation does not depict “expected” effects to be measured in the forthcoming trial that will determine Payments to Village Enterprise. One reason is that the distribution is sensitive to specification and operationalization choices discussed in the main paper, and that these may change in a future evaluation.

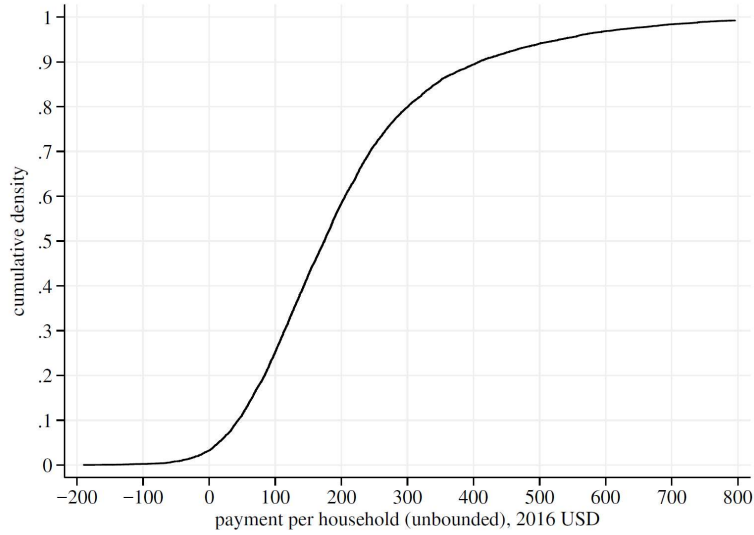
Note: All monetary values, including those used as calculation inputs, are defined as current 2016 dollars.

Figure 4: Simulated Payments P_m , Unbounded



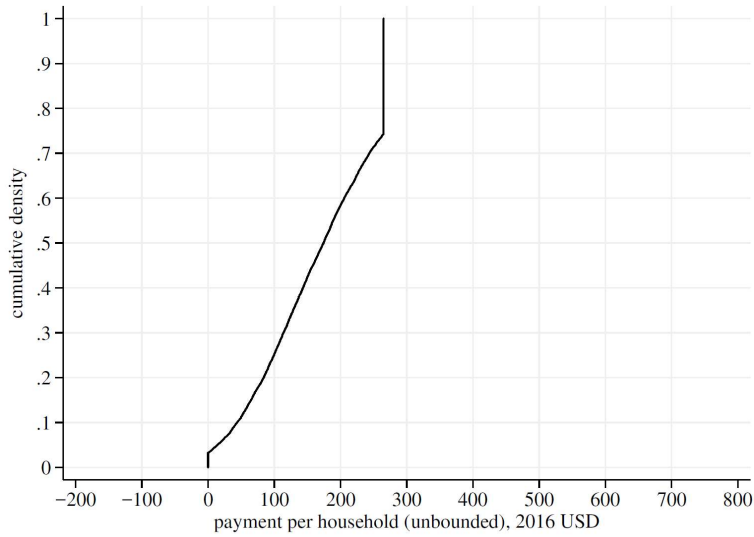
Explanation: This depicts simulated 5,000 payments that would have been associated with each of effects depicted in Figure 3. It assumes that equations (VI) and (X) apply for all observations (i.e., no adjustment for scenarios A and B); a period m of 21 months; a long-run discount rate 10%.

Figure 5: Simulated Payments P_m , Subject to Scenario Bounds



Explanation: This equals Figure 4, with the edit that growth rates are bounded by scenarios and associated equations (VII), (VIII), and (X). Note that unlike in Figure 4, more than 99% of payments are now within the observable range on the x axis.

Figure 6: Simulated Payments P_m , Subject to Scenario Bounds and Payment Bounds



Explanation: This equals Figure 5, with payment caps and floors integrated.