

**Social and Environmental Crediting:
An Impacts-Based Strategy for Efficient and Sustainable Development Funding**

Robert Van Buskirk and Benjamin P. Eifert¹

This draft: August 14, 2008

Abstract

This paper describes the notion of *development credits* generated by verified social and environmental impacts in much the same way that carbon credits are generated by verified reductions in carbon emissions. In this approach to development financing, donors set prices for credible evidence of progress towards specific goals like improved standards of living, reduced disease burdens or improved environmental conditions in poor countries. This generates a direct link between resource flows and clear scientific evidence, generated by independent rigorous evaluations and audited output statements, of the incremental impacts of programs run by NGOs, governments, communities and businesses. To give an example, an NGO running an anti-malaria program that directly resulted in an estimated ten thousand fewer Nigerian children dying from malaria would earn development credits valued at ten thousand times the current price offered by donors for saving a child's life. The most ambitious version of this approach is a global market for credits generated by verified development impacts, allowing the most effective organizations and programs to finance their operations through impersonal market mechanisms that measure and reward performance.

A development credits system would bring important advantages, including shifts in the allocation of development finance towards more cost-effective programs and organizations, stronger incentives for innovation and for good management, less centralization, reduced transactions costs, and ultimately more impact for each dollar spent. However, such a system, especially in its most ambitious form, requires a strong institutional framework capable of providing credible and transparent information about program impacts, efficient prices for impacts, fraud deterrence and mitigation of unintended negative consequences and perverse incentives. This paper examines these advantages and challenges in the hope of generating a constructive policy dialogue.

¹ Lawrence Berkeley Labs and Department of Economics, University of California, Berkeley.

Table of Contents

1	Introduction.....	3
2	Motivation.....	8
2.1	Big returns, big challenges	8
2.2	Carbon Markets.....	11
2.3	Applying these ideas to development.....	15
3	The Case for a Market for Verified Social and Environmental Impacts.....	17
3.1	Efficiency of resource allocation	18
3.2	Incentives for innovation and better management.....	20
3.3	Decentralization and empowerment	21
3.4	Summary and remarks.....	22
4	Making It Work	22
4.1	Outcomes and market segments.....	23
4.2	Verification of program impacts	24
4.3	Translating impact evaluations into development credits.....	30
4.4	Pricing mechanisms	32
4.5	Perverse incentives and unintended consequences.....	35
5	Summary and Conclusion.....	40
6	References.....	42

1 Introduction

The global effort to reduce absolute poverty and improve human conditions has gained in visibility and importance in recent years. The idea that everyone should be entitled to basic economic, social and environmental rights regardless of race, gender or citizenship, and that the international community should work together to achieve those basic standards as fast as is reasonably possible, has become widely accepted. In 2000 the international community reached a consensus around the eight Millennium Development Goals (MDGs):

1. Eradicate extreme poverty and hunger;
2. Achieve universal primary education;
3. Promote gender equality and empower women;
4. Reduce child mortality;
5. Improve maternal health;
6. Combat HIV/AIDS, malaria and other diseases;
7. Ensure environmental sustainability;
8. Develop a global partnership for development.

Most of these goals relate to concrete, measurable improvements in human conditions.. The challenge faced by the development community is how to accelerate progress on these goals.

One mechanism for doing so is to increase the financial resources available to aid agencies, an approach championed in the academic and policy community by Jeffrey Sachs among others.² This is an important step, but as many economists and development practitioners note, it is unlikely to achieve the hoped-for results without significant corresponding improvements in the quality of aid and the efficiency of its allocation. It is hard to dispute that much aid over the past four decades has achieved disappointing results in poor countries.³

Improving aid effectiveness requires better mechanisms for selecting high-return projects and ensuring high-quality implementation. Over recent years, nongovernmental organizations, foundations and aid agencies have increasingly emphasized goals, targets, monitoring and

² See Sachs (2005), *The End of Poverty*. Penguin press.

³ See the references in Collier (2007), Collier and Gunning (1999), Ndulu et al (2007), and Easterly (2003). Some cross-country studies have found positive correlations between aid flows and economic growth in particular types of countries, but Roodman (2007) demonstrates the lack of robustness of these findings.

evaluation. This is part of a broad-based attempt to measure the effectiveness of projects in reducing poverty and improving health and educational outcomes, and to channel resources towards those activities that can have the greatest impact. However, the mechanisms linking credible measurement of program impacts to the allocation of financial resources remain tenuous.

Some existing proposals argue for directly linking foreign aid to demonstrable progress against agreed country-wide goals. This is often called “output-based” or “progress-based” aid. For instance, Birdsall, Savedoff and Vyborny (2007) propose an experiment in which donors offer low-income countries \$100 per child who completes primary school and scores above a baseline on a standardized competency test. This approach improves incentives for developing country governments, who receive more aid money the faster their programs and policies translate into reduced poverty rates and improved educational and health outcomes.

Here we offer a different yet complementary approach, one that offers a unique ability to decentralize information and decision-making, strengthening incentives and opportunities throughout the organizations involved in designing and implementing development programs. Our vision is for a crediting system in which organizations which make credibly and objectively verified progress on agreed social and environmental goals in developing countries receive credits in proportion to that progress. Such credits can then be sold to donors or concerned citizens from rich countries in an impersonal market. The market institutions set out the outcomes for which credits are assigned: increased consumption for people below the poverty line, reduced child mortality, reduced disease burden (e.g. measured in disability-adjusted life-years, or DALYs), increased primary school completion, accelerated reforestation, and so forth. Different credit types or market segments would be defined accordingly. Every program presenting evidence from an independent evaluation meeting the system’s standards would be eligible to apply for credits corresponding to the type and magnitude of its impacts. These credits would be priced as a function of the supply of credibly evaluated programs and the demand for development progress from donors and rich country citizens. The idea is simple: if it can be independently and rigorously demonstrated that your program increases the health of your target population by one million disability-adjusted life-years (DALYs), you get paid the market price for one million DALYs.⁴

The potential advantages of this approach over more traditional, centralized development finance allocation mechanisms are substantial. By creating a tight link between financial flows and results on the ground, this approach channels available resources to the organizations and causes which

⁴ With qualifications discussed below, including adjustment for the statistical certainty of the evaluation.

create the most social value for every dollar spent. That same link creates strong incentives for organizations to design better projects and implement them more effectively, thus stimulating innovation and incentivizing better management. The results-based approach is inherently decentralized, empowering local actors and providing an avenue for experimentation and new approaches. This helps overcome the problem that innovative approaches tend to be initially under-funded, creating barriers to innovation, and then over-funded once they gain popularity, reducing quality control and long-run effectiveness. A market for verified development impacts also allows a wider variety of donors with limited capacity for rigorous monitoring and evaluation to know what impact their charitable dollars are having, potentially raising the supply of funds available. We present these arguments in more detail below.

Overcoming the market failures that prevent the natural emergence of this type of market will require a strong institutional framework. First, it requires a set of agreed-upon social and environmental outcome indicators that define the market itself. These should be simple, easily measurable and parallel the global consensus as reflected in the MDGs. Second, it requires credible verification of the impacts of individual programs on meeting consensus goals. We argue that rigorous evaluations, which provide unbiased estimates of program impacts using scientific, peer-reviewed methods, must be central to a market-based approach to development financing, though not every program need have a full randomized trial (as discussed below). If the international community is indeed to pay for results, developing a system to monitor program impacts through rigorous evaluations and to “evaluate the evaluations” is absolutely crucial. Finally, the system requires a mechanism to price verified development impacts. Fourth, it must have built-in safeguards to mitigate unintended negative consequences of an incentive-based approach, including increased liquidity constraints and risk exposure for implementing organizations like NGOs, as well as any potential for gaming and cheating. This institutional framework would be anchored around a sophisticated market operator that sets, monitors and enforces clear standards. This role could be played by an individual donor agency setting up its own crediting program or by an independent market operator aiming to create a global marketplace for verified development impacts.

The tools of rigorous program evaluation are central to making such a system work efficiently. Rigorous evaluation is the science of correctly determining the counterfactual for an intervention – that is, what outcomes in the target population would have been in the absence of the intervention – and hence estimating its true causal impact. It may be infeasible for every project to have a full-blown randomized impact evaluation that provides estimates of direct effects on the final outcomes we care about, like consumption, mortality, morbidity or biodiversity. However, global knowledge

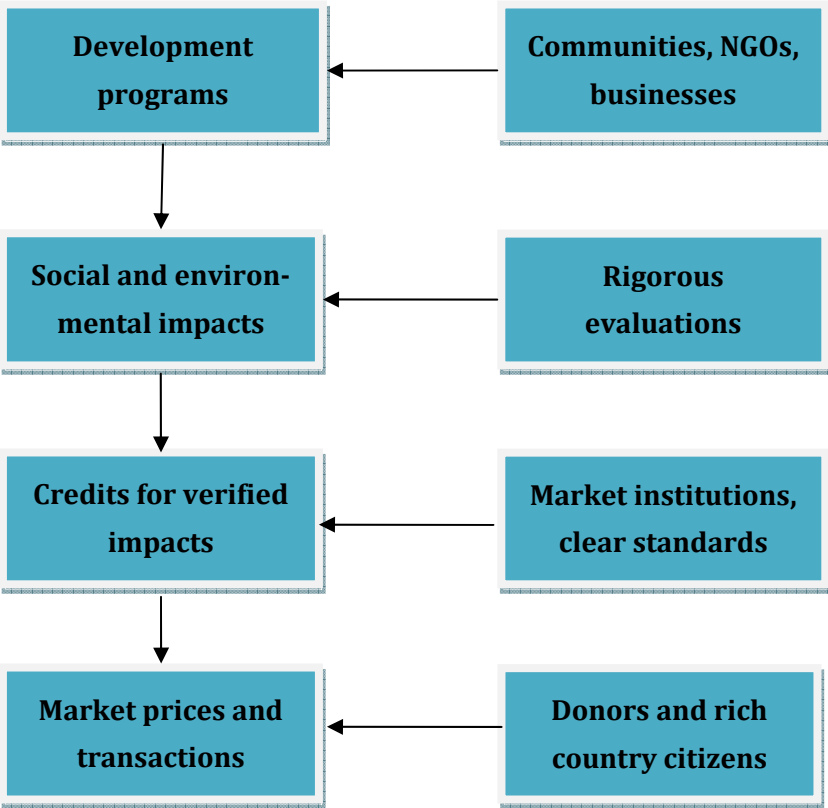
from existing rigorous impact evaluations can help the crediting system link these final outcomes with more easily measurable outputs like jobs created, increased vaccination rates or new trees planted in a deforested area. The system could price verified outputs at a discount relative to outcomes, with the discount adjusted for the level of certainty with which the relationship between the outputs and outcomes in question is known. The prices paid for verified improvements in intermediate outcomes could be recalibrated as new impact evaluations improve the state of our knowledge of their relationship to final outcomes.

We recognize that a full-blown global market in verified development impacts is a long-run vision indeed. However, the principles of the approach outlined here can be implemented on any scale and evolve according to their successes and failures. Over time a crediting system set up by a consortium of donors could grow as its success attracts new participants, eventually beginning to look more like modern commodities or carbon market.

Even in the long run, a market for verified development impacts is likely to be a complement, rather than a substitute, for other forms of donor engagement. While foundations might direct some of their resources into such a market, they also retain an important role in screening and providing seed funding to new social entrepreneurs and upstart NGOs, enabling them to make up-front investments in high-value programs which will generate development credits down the road. Some development programs may provide important services but have difficulty quantifying their impacts; foundations and aid agencies can continue to play a role in funding such programs. Donors also would continue to provide emergency aid in crisis settings that are deemed too dangerous and disorganized for proper monitoring and evaluation but too critical to ignore.

This discussion paper is organized as follows. Section 2 offers some motivation for our proposal, discussing existing aid financing mechanisms and recently developed markets for carbon emissions reductions. Section 3 discusses the argument for a social and environmental crediting system, holding aside the institutional design issues necessary to make it work. Section 4 raises and addresses these practical issues in more detail, illustrating the kinds of institutions and systems that might work but not claiming to have all the right answers. Section 5 turns to residual concerns raised by an incentives-based approach such as gaming and unintended consequences and mentions how some of these challenges might be addressed. Section 6 concludes.

Figure 1. The Development Marketplace



2 Motivation

The tragedy of absolute poverty around the world needs little introduction here. There is relatively little disagreement about the end goals of development assistance, for instance as they are articulated in the Millennium Development Goals. The purpose of this discussion paper is to suggest a new type of mechanism for accelerating progress on these goals.

This section briefly discusses the background of development aid and the challenge of increasing its effectiveness, provides an overview of the carbon markets from which the proposals here draw much of their motivation, and then discusses the potential applicability of the core ideas on which carbon markets are based to broader social and environmental problems in developing countries.

2.1 Big returns, big challenges

Developing countries offer many opportunities for generating high social returns at low cost. To give one example, rural African households often spend 200-300 hours per year collecting fuel for cooking while urban households may spend up to 20% of household income on energy. Improved cook-stoves in Rwanda have been reported to save 394 kg of charcoal per year worth \$84 for a cost of only \$12 per stove (Barnes, et al, 1994). In another example, children in rural areas with poor sanitation infrastructure often suffer from parasites which sap their energy and lower their school attendance rates, while de-worming drugs that cost less than \$0.05 per dose administered every six months are highly effective in eliminating these parasites and slowing their transmission among school-age children (Miguel and Kremer, 2005).

These high-return opportunities provide ample motivation for commentators calling for increases in aid flows. Over 1990-2005, ODA flows totaled \$111 billion to Latin America and the Caribbean, \$115 billion to South Asia, and \$358 billion to sub-Saharan Africa. The total to all low- and middle-income countries was \$1.25 trillion.⁵ The international consensus around the MDGs involved a promise of doubling (or tripling, in Africa's case) the magnitude of aid flows moving forward. However, as Barder and Birdsall (2006) note:

⁵ World Bank, *World Development Indicators 2006*.

There are significant differences of opinion about the merits of additional aid in meeting the MDGs, including whether and how aid should be given in 'fragile states', whether additional aid on the scale envisioned can be effectively used even in well-managed economies, and whether the aid system, particularly in highly aid-dependent countries, undermines instead of strengthening local institutions.⁶

Simply put, money may not do very much without better execution. The challenge is to link resources to organizations with the capability to deliver on high-return opportunities efficiently and at adequate scale, often in difficult environments. To put it mildly, this is hard. Governments have the scale and scope to coordinate implementation of large projects, and historically the vast majority of aid resources have flowed through them. However, developing country governments often have weak capabilities and poor internal controls, and ministries are often highly centralized and disconnected from local needs. Oversight by international institutions has thus far demonstrated limited ability to improve the internal functioning of governments in weak states.⁷

The limited capability and reach of many developing-country governments has spurred an explosion of involvement by nongovernmental organizations (NGOs) and community-based organizations (CBOs) in development projects around the world over the last few decades. NGOs and CBOs both raise resources to design their own projects and serve as implementing agents for projects commissioned by donor agencies. Such organizations often have better capacity than governments to implement specific projects and work with local communities. However, their scale and scope (both substantive and geographical) is often limited, and it is difficult for donors or governments to coordinate and monitor the activities of many small service providers.

A common limitation of all types of development assistance is that it is difficult to know with reasonable accuracy the impacts that funds are generating. Governments in very poor countries often have weak internal budgeting and monitoring systems, to the point that it can be difficult to ascertain how money was spent.⁸ Aid money is also inherently fungible.⁹ Empirical research in macroeconomics has tried to estimate the average impact of aid on economic growth and on other social outcomes, but this task is arguably difficult or impossible with the existing data.¹⁰ This leaves donors with their qualitative judgments on how effectively their direct-to-government aid is being used and what impacts it is having, but often little more than that.

⁶ Barder and Birdsall (2006), page 3.

⁷ See, for instance, Dichter (2003).

⁸ See in particular the World Bank's Public Financial Management Assessments.

⁹ See Pack & Pack (1993), van de Walle & Mu (2007).

¹⁰ See Rajan and Subramanian (2007).

The idea of *output-based aid* to governments is a promising response to these challenges. The basic idea is akin to the broader notion of selectivity, or allocating aid flows based on the demonstrated past performance of governments.¹¹ The key distinction is that output-based aid is assessed based on verified outputs like poverty reduction and improved educational attainment, rather than on process inputs like ratings of institutional quality or adoption of a certain set of policies. A recent proposal by Owen Barder and Nancy Birdsall (2006) argues for this:

*The approach, simply stated, is to link additional aid to clear evidence of progress already achieved on the ground. The approach would give flexibility and autonomy to local institutions, providing an opening for local institutional experimentation, while at the same time ensuring that aid pays only for real, measurable achievements.*¹²

Output-based aid has its drawbacks,¹³ but it provides simple and clear incentives for governments: build effective programs which generate bottom-line impacts, and the aid will flow. Given how difficult it is for donors to know the true quality of government policies and programs, rewarding outputs rather than effort is an attractive idea.

Similar trends have taken hold in the non-governmental space as well. Without scientific evidence on program impacts, donors must rely on qualitative judgments about NGOs, their staff, their program designs and their implementation work when deciding which organizations and programs to fund. However, in recent years many NGOs have partnered with economists and evaluation experts to bring rigorous impact evaluation methods into their projects, encouraged by high-visibility donors like the Gates Foundation and the World Bank. Rigorous evaluations provide

¹¹ For instance, the World Bank now uses a comprehensive Country Policy and Institutional Assessment (CPIA) to rate governments' performance on economic policy, poverty reduction, service provision, corruption control, environmental protection, gender equality, and so forth. It then determines the amount of aid available for each country via rigid formula based on its CPIA score. The US Millennium Challenge Account (MCA) has a similar logic, screening countries on the basis of several criteria related to governance and only offering resources to those above a certain threshold.

¹² Barder and Birdsall (2006), page 3.

¹³ Outputs depend on the quality of government policies and programs, but they also depend on many other factors, including sheer randomness. Hence output-based aid depends in part on things which are outside of the control of governments, which generates excess volatility in aid flows. As an example, consider what the AIDS epidemic would have done to aid flows in southern Africa if aid had been linked heavily to health outcomes as early as 1990. If donors could perfectly observe governments' actions – e.g. how aid money is spent and how well projects are implemented – and knew the exact relationship between those actions and the social outcomes we care about, then the best approach would be to instead tie aid flows to those actions rather than to outcomes. However, in conditions of poor information about the quality of government actions and the relationship between those actions and social outcomes, the second-best approach may be to focus aid flows on rewarding demonstrated success rather than trying to quantify and reward effort.

scientific evidence on impact and cost-effectiveness, helping donors to learn about what types of interventions to fund and helping high-quality NGOs build evidence-based reputations. The movement towards rigorous evaluation offers an important complement to output-based aid to governments, by providing better information and incentives to the large array of private funders and service providers working in developing countries. To date such evaluations are carried out in only a small minority of development projects, but the trend is moving steadfastly in the right direction.

We view our arguments in this discussion paper as a natural, if ambitious, extension of these trends. If good-quality information about program impacts is available, then a market which puts a price on verified social and environmental impacts would create strong incentives for public and private organizations to identify opportunities for generating large impacts at low cost. This is very similar to the basic idea underlying carbon markets.

2.2 Carbon Markets

The global fight against poverty and environmental degradation has in its scale and complexity many elements in common with the international effort to reduce carbon emissions. In both cases, achieving a certain shared goal depends not only on allocating adequate resources but also on influencing the actions of thousands or millions of individuals and organizations across the world in a way that maximizes the impact of the available resources.¹⁴

The basic idea

Markets in tradable permits for carbon emissions were originally based on ideas developed in the 1930s by economists associated with the Technocracy movement.¹⁵ The idea is to minimize the cost of preventing or reducing the extent of climate change by using policy instruments which ensure that the firms which reduce their emissions the most are those with the lowest costs of doing so. In a standard cap-and-trade carbon market, the market operator gives out a fixed quantity of permits that entitle firms to emit a certain amount of carbon and allows firms to buy and sell those permits

¹⁴ One substantial difference is that externalities are more central to the issue of carbon emissions. Individuals and businesses bear the costs of reducing their own emissions, while future generations across the whole world benefit from emissions reductions anywhere in the world. One can make a case for the existence of externalities in the case of development – for instance, if extreme poverty and inequality increases the probability of conflict and terrorism that affects rich countries – but it is more indirect than in the case of climate change.

¹⁵ See Akin (1977).

freely.¹⁶ Those who do not fully use their entitlement can trade the surplus emissions reductions with those who cannot cost-effectively meet the emissions reductions implied by their entitlement. As a result, firms which can reduce their emissions cheaply do so and sell their permits to firms which cannot, allowing emissions to be reduced by a given proportion for the least overall social cost. The cap, or quantity of permits available, can be lowered over time in accordance with overall emissions reduction targets. This type of scheme is also used in a variety of regulatory regimes for other pollutants like oxides of sulfur, mercury and nitrogen oxides.¹⁷

Under the Kyoto Protocol's Clean Development Mechanism (CDM), industrialized countries with greenhouse gas reductions commitments can fulfill those commitments in part by financing programs in developing countries, potentially at much lower cost than would be possible domestically. As a result, an NGO or government agency in Nigeria can implement an innovative biomass fuel efficiency program, apply to CDM and go through the verification process, and then receive funding in proportion to the impact of that program on carbon emissions. This creates powerful incentives for organizations to seek out and implement high-yield, low-cost emissions reductions programs in developing countries. Some private business enterprises have been started to focus exclusively on the CDM market, and more than thirty specialized service providers have entered since 2005 to help organizations and firms in developing countries measure emissions, define strategies for emissions reductions, and connect to global carbon markets. CDM accounted for 49% of the total volume in the carbon markets in 2005, with Certified Emissions Reductions (CERs) worth \$11 billion (Harris 2006). By 2006 the value of CDM credits had climbed to \$30 billion, corresponding to about 450 million tons of CO₂ equivalent. As long as the verification and auditing systems work properly in determining true program impacts, and as long as transaction costs and barriers to entry are not too high, the market aligns net social welfare creation with private profits.

Voluntary offset markets have also emerged to supplement their larger Kyoto cousins. These markets are home to smaller projects, averaging around 5,000 tons CO₂ equivalent per year and totaling around \$100 million in 2006, which cannot overcome the transactions costs and verification standards for CDM. The voluntary markets provide a niche for projects which have greater uncertainty over actual emissions reductions and for those rich-country donors who have relatively high tolerance for that uncertainty.

¹⁶ An alternative to the entitlement approach is to require firms to bid in an auction for emissions permits. This approach is rarely used in practice, though more for political reasons.

¹⁷ See Teitenberg and Johnstone (2004) for a thorough survey of the literature on tradable permits.

Practical challenges

While their potential for lowering the cost of reducing emissions is substantial, carbon markets face substantial challenges in practical implementation, and critics have pointed out the importance of several of these in the Kyoto-CDM and the voluntary frameworks in place today. Key issues include the measurement and verification of emissions reductions, the permanence of reductions, whether or not emissions reductions are *additional*, the ability of firms to cheat the system, and whether or not to try to capture some of the high profits accruing to low-cost projects.

First and foremost is the issue of measurement and verification of impacts. The scientific standards required for verification of impacts with a high degree of certainty are stringent and hence expensive and burdensome to implement. The transactions costs of getting a CDM project through are very high, potentially millions of dollars, and account for 20-40% of the total value of certified emissions reductions (CERs) generated by most CDM projects; see Harris (2006). This means CDM is only accessible to large projects. Smaller projects can target the voluntary credits market, where the average project size is around 5,000 tons of carbon equivalent, compared to more than 50,000 tons for CDM projects. Labels assigned by standards organizations can help provide quality control in the voluntary market. For instance, the Voluntary Gold Standard requires all projects to have their technical datasheets and verification plans validated by a UN-accredited auditor, and then uses contributions from all labeled projects to audits a random sample of projects. However, the majority of credits on the voluntary market remain unlabeled. As a result, the real impacts of projects selling credits on the voluntary market are not very transparent, and prices paid for these credits vary a great deal, from €0.10 per ton of carbon equivalent to more than €50 per ton in the year 2005 (Harris, 2006). In comparison, prices on the CDM market ranged from €8.5 to €20 per ton in 2005.

A related point is the issue of permanence of impacts. For example, a forestry project may claim an impact corresponding to the carbon-reducing effect of the new trees over a long time horizon, but if the trees are later cut down or burned in a forest fire then the true impact of the program is much smaller. Taking these types of uncertainty into account in estimating program impacts is difficult but crucial. In fact, several standards organizations for voluntary carbon markets discount credits generated by forestry projects heavily, require a deposit of a some credits into a reserve fund, or refuse to label them altogether.

In addition to measuring impacts, certification of carbon emission reductions also usually requires demonstrating *additionality*, or that the program would not have been viable without the revenues

generated by the sale of carbon credits. This is difficult to verify in practice, and the CDM has interpreted the additionality requirement as a mandate that 10% of projected project revenue comes from carbon credits.

Another concern is the possibility that organizations might try to sell the same emissions reductions multiple times on different markets. Most operators in the voluntary market have their own credit registries, but these are not currently inter-operable, so this type of gaming is potentially possible until technical cooperation improves.

A final, somewhat tricky question is whether it makes sense to try to capture some of the high profits from easily-identified low-cost projects by some other mechanism. The carbon markets are designed to incentivize cost-effective carbon-reduction projects by offering a fixed price per verified ton of carbon equivalent. If a project has very low implementation costs, then that project receives a large surplus or profit from the carbon markets. For example, by 2012 an estimated €4.7 billion in credits will have been granted for low-cost reductions in HFC-23 emissions by firms producing refrigerants gases. Critics argue that this surplus reflects a waste of public resources, and that a better policy would be to pay for the technology required to capture and destroy HFC-23 in developing country industries, which might cost as little as €100 million (Wara, 2007).¹⁸

In some ways this is correct: in a world of perfect information and perfectly effective bureaucracies, centralized allocation and funding of projects is the cheapest way for the public sector to elicit reduced emissions. The problem is that central policymakers are ill-equipped to perform the task of analyzing the vast array of potential carbon-reducing projects throughout the world, identifying those which are truly most cost-effective and overseeing the efficient implementation of those projects, for all the same reasons that they are ill-equipped to micro-manage the economy in the same way. In principle, if there exists an easily-identified set of obvious, low-cost, easy-to-

¹⁸ Wara also argues that the presence of these cheap sources of credits reduce the incentives of major carbon producers like power utilities in China and India to install more expensive but lower-emitting electricity generation facilities, which will be the major driver of reductions in carbon emissions over the medium to long term. This is true to the extent that the existence of very-low-cost projects lowers market prices for carbon credits. However, if these very-low-cost projects are exhausted after five or ten years, market prices will rise in the future, which should provide adequate incentives for forward-looking investors sinking big money into long-lived power plants. If, on the other hand, there will be a sufficiently large supply of very-low-cost projects to allow overall medium- to long-term emissions reductions targets to be met with low carbon prices, then higher-cost efficiency improvements in power plants are unnecessary and undesirable. The exception to this argument would be if carbon markets are mispricing different types of emissions (CO₂, nitrous oxide, methane, HFC-23, etc.) with respect to their long-term impact on climate change, in which case there is a strong argument for fixing the crediting rules or separating the overall market into smaller markets for each category of emissions, as Wara suggests.

implement emissions reductions projects, it might make sense to deal with them in a separate framework while allowing the market to operate for other projects. However, this is a slippery road to go down, as over time the market will identify new low-cost types of projects, and policymakers will be tempted to “skim” them off the top as well.

2.3 Applying these ideas to development

Carbon credit markets are built around a system that (i) defines a global goal, reducing greenhouse gas concentrations in the atmosphere; (ii) allocates entitlements or property rights, allowable emissions levels for countries under the Kyoto protocol or under CDM baseline emissions projections; and (iii) creates a market in which those property rights can be monetized, creating incentives for efficient allocation of resources. By introducing property rights over emissions reductions and pricing those property rights, we create incentives for emissions reductions to be carried out by those who are best able to do so at low cost.

In contrast to mandatory emissions reductions, the goals of the international effort to reduce poverty and reverse environmental degradation are “soft” goals which cannot be mandated through caps or targets, but must be achieved through incremental progress over time. Nonetheless, similar principles can be applied. To give an example, consider the two programs designed to improve educational attainment in Kenya discussed by Kremer (2003). The first is a class size reduction program, a popular instrument used in developed countries based on the notion that teachers are more effective when they can give students more individualized attention. The second is a textbooks program aiming to give students better resources with which to learn on their own and in conjunction with classroom instruction. Both are well-intentioned and their ideas are certified by educational experts. But rigorous evaluations demonstrate that the textbooks program improved student outcomes thirty times more than the class size reduction program for each dollar spent. Given a budget for funding educational programs in similar villages in Kenya, one way to improve outcomes the most rapidly for the most people is to allocate more money to textbook programs and less money to class size reduction programs. By shifting resources to more cost-effective programs we increase the total social impact generated by any given quantity of development financing.¹⁹

¹⁹ The underlying logic is that if (a) donors define their goals in terms of human entitlements, (b) such entitlements are universal and non-discriminatory, (c) incremental progress towards entitlement satisfaction can be quantified, and (d) the overall goal is to satisfy these entitlements within the shortest possible time with limited resources, then it follows that projects that can demonstrate more cost effective delivery of progress towards entitlement satisfaction should be allowed to get “at the front of the line” for resource and budget allocations from donors. The overall level of funding that donors set then implicitly defines a price for incremental improvements in social outcomes that donors are willing to pay.

The question is, how can the international community shift a greater proportion of aid money into high-impact activities? One starting point is encouraging greater use of the tools of rigorous impact evaluation throughout the development community, so that impacts and cost-effectiveness can be compared across a wide range of programs. Carbon markets offer an example of how this can be taken a step further. If a market system exists to price social and environmental impacts in developing countries, and if a credible evaluation system exists to verify those impacts, this will create powerful incentives for organizations to seek out and implement high-impact, low-cost programs, and resources will flow in response. These potential benefits are the subject of Section 3. However, as with carbon markets, getting the institutional details right will be difficult (but we think, possible); this is the subject of Section 4.

3 The Case for a Market for Verified Social and Environmental Impacts

Markets are powerful phenomena when they work well. In Silicon Valley, every year hundreds of engineers and entrepreneurs design innovative new technologies and start businesses to apply those technologies to real-world problems. Their work, while uncertain and risky a priori, has clear, objective performance benchmarks which link the financial rewards they earn to the long-run value created by the companies they start and the products and services they sell. Venture capitalists seek out the people with the best ideas and the best teams and invest money and time helping them to reach their potential. In this way, resources flow to the highest-value projects, actors have every incentive to design the best projects and invest in high-quality management and implementation. In addition, investment decision-making is decentralized to the people and organizations which are closest to the technologies, consumer needs and local economic conditions. This all works because performance is objectively and credibly measurable and directly rewarded, while those with impractical or infeasible ideas ultimately face a declining or stagnating business or operation.

Our vision is a market in which donors and individuals can purchase credits generated by programs which make credibly and independently verified progress on a set of core social and environmental goals. The market institutions set out the basic outcomes in which donors are interested – increased consumption for people below the poverty line, reduced child mortality, reduced disease burden (e.g. measured in DALYs), increased primary school completion, accelerated reforestation, and so forth – and different market segments and credit types are defined accordingly. Any program presenting evidence from an independent evaluation that meets the standards set out by the market operator is eligible for credits corresponding to the type and magnitude of its impacts, which can be sold impersonally on the market for a price determined by supply and demand. As a result, opportunities for generating high social returns at low cost are sought out, innovation is rewarded, resources flow to programs which improve poor people's lives, and donors of any size and sophistication know the impact that their funds generate.

Implementing this vision requires overcoming many serious market imperfections, first and foremost the scarcity of credible information on program impacts. In this section, we present our arguments on the advantages of a market for development credits in accelerating progress on the state of human development. We ask the reader to evaluate these arguments assuming initially that

the systems necessary to make it work can be built. The challenging issues regarding how to build properly functioning market institutions are the subject of Section 4.

3.1 Efficiency of resource allocation

In the standard approach to project finance, donors (foundations, governments, aid agencies, and so forth) consider project proposals from different teams or organizations and decide which to fund. Some donors are quite sophisticated and use tight screening methods: they evaluate the ideas, the objectives, the implementation plan, and the quality of the team involved, often benchmarking against past experience in the relevant countries or sectors. However, in the end the central officials who make final decisions about funding have very limited information about what the impacts of the proposed programs will be, reducing the efficiency of resource allocation.

Imperfect information also creates an important incentives problem in the centralized allocation process. The best available information about the potential impacts of a proposed intervention usually lies with the implementing team, particularly if its members are experienced and have done adequate background work. This generates what economists call *agency problems*. The donor (or the principal) would like the implementing team to give a completely frank, objective assessment of likely impacts and downside risks. However, the implementing team (or the agent) is prone to over-optimism: it is competing against other proposals for the same pot of funding and it knows that donors want to fund high-impact programs. On the margin, implementing agents face a real trade-off between spending more effort uncovering private information about the potential impacts of their next project on the one hand and deepening their relationships with donors and producing compelling presentations on the other. NGO officers complain vigorously about the effort they have to spend chasing money and courting donors in the current funding environment. In the end, presentation skills, packaging and connections to donors often play a significant role in who gets a grant and who does not. This can lead to an over-investment in time and effort spent fundraising and an under-investment in implementation.

Reputation effects, trust and repeated interactions can help alleviate agency problems, especially if programs are rigorously evaluated, as donors can identify and avoid NGOs and governments which regularly over-sell their work. This is the basic idea behind GiveWell.net, which attempts to publicize the effectiveness of charitable activities in order to guide resources to high-impact uses. However, such efforts are limited by the scarcity of rigorous evaluations. This is due in part to the newness of the methods and a lack of capacity in many organizations to implement them, but also

to the strong externalities inherent in knowledge creation. Standard audits and procedural checks, the main alternatives to rigorous evaluation, can detect very bad outcomes like failure to deliver agreed-upon inputs or outright fraud, but cannot help donors discriminate meaningfully between programs that meet their basic procedural standards. This leaves donors to make decisions on the basis of lower-quality information than is ideal.

In a market for development credits, NGOs and other implementing agents receive tradable credits in proportion to the demonstrated results of their programs on well-defined outcomes. The net financial flow to an implementing agent from a completed program is its social value minus its cost, so even if an implementing organization is solely concerned with the financial viability of its operations, it will still effectively make decisions on the basis of expected net social welfare. Programs which are likely to create a lot of social value for low cost will be implemented, whether or not their originators have connections in the donor community or their ideas resonate with the latest fads in development practice. Ideas which sound great but in practice are likely to generate few demonstrable results will likely be dropped.

The market has the important effect of aligning information and decision-making authority. A program will be launched if its implementing agents – the people who are the best informed about its true potential and its downside risks – are willing to stake their own resources on the belief that they will create social value that exceeds its costs.²⁰ Since the implementing agent or project investor pays the operational costs of a program and earns development credits based on its results, it has every incentive to uncover the best information about the program's likely impacts and no incentive to over-sell those impacts. While implementing agents have imperfect knowledge of the results their potential programs will bring, the market system rewards relevant experience and generates strong rewards for objective analysis of potential program impacts. In this environment, large-scale rigorous evaluations act as public goods, making forecasts of the potential impacts of similar programs more precise and cheaper to produce.

As implementing agents launch programs which offer them the best financial outlook, aid money will flow from programs with lower expected social value per dollar to those generating the most bang for the buck. As a result, the overall impact of the portfolio of development programs will rise, and the cost of saving lives and improving livelihoods in poor countries may fall. This logic is akin to the functioning of a market economy as compared to a command economy: as much as we might

²⁰ This is somewhat loose because implementing agents are likely to be risk averse and face credit constraints; more on this below.

believe in the intelligence and analytical capabilities of experts making centralized resource allocation decisions, often it is better to decentralize decision-making to agents who have on-the-ground information by using incentives and prices.

A useful byproduct of the information system which allows the market to function is that donors can accurately estimate the efficiency and social and environmental impacts of a subset of their overall funds and benchmark their impacts against those of other donors. This would provide foundations and aid agencies with a reality check on their activities and facilitate some healthy competition.

The discussion above glosses over several important subtleties. Most importantly, payment for ex-post realized impacts means that implementing agents must still secure seed funding for their work up front. This would make NGOs operate more like private businesses, raising working capital initially to fund investments in programs and recouping costs by selling the credits earned, and would expose them to potentially significant financial risks. On one hand, this creates important incentives for them to design and launch programs with high expected social returns and to manage them as well as possible. On the other hand, NGOs and community organizations often lack financial depth and are not well positioned to absorb substantial financial losses from programs that turn out to be less effective than anticipated. This suggests the importance of a fostering a specialized financial services market to support the operations of NGOs focused on the development credit marketplace. It also suggests a role for foundations as “venture philanthropists” which identify and seed promising NGOs and programs, with the markets providing a sustainable source of financing to those which are successful. We discuss this more in Section 4.

3.2 Incentives for innovation and better management

In the private sector, firms are rewarded for creating new products and developing better processes by reaping the benefits of increased sales and higher margins. Markets provide incentives to take intelligent, calculated risks to stay ahead of competitors (Gilbert, 2006), and the evidence suggests that firms in reasonably competitive product markets innovate more (e.g. Blundell, Griffith and Van Reenen 1999; Bloom et al (2005). Competitive pressures also push firms to find people and technologies that get things done quickly and well. New empirical research finds strong correlations between the quality of management practices in manufacturing firms and their rate of innovation on the one hand and the competitiveness of their market environment on the other (Bloom et al, 2007).

Innovation and execution are equally important in the social sector. The notion of “social entrepreneurship” emphasizes creativity, sound analysis and high-quality management practices in nongovernmental and community-based organizations. Market pressures and incentives can accelerate this trend of organizational effectiveness. A market for verified development impacts facilitates innovation and incentivizes better management because implementing organizations are effectively paid for successfully improving the quality of their programs. The objective focus on results rewards well-thought-out experimentation: an NGO or government agency which believes strongly in an innovative project concept that runs counter to established development thinking is much more likely to execute it if it can sell development credits down the road than if it must convince skeptical donors in advance. The market also leads implementing agents to discriminate between truly effective management tools and promising new ideas on the one hand and gimmicky fads on the other.

3.3 Decentralization and empowerment

The discussion above already mentions the way markets move decision-making downward from centralized officials to the people with the best information. This has a powerful decentralizing effect. With a market for verified development impacts and credible small-scale evaluation methods, community-based organizations with low costs and strong local knowledge can access the funds to begin to play a larger role in development practice. The existence of the market would likely lead to the emergence of NGOs and businesses which specialize in finding promising community-based organizations and providing them with the technical assistance necessary to validate their programs and generate development credits. This has already been seen in the carbon markets, with more than thirty specialized for-profit and non-profit service providers emerging since 2005 (Harris, 2006). Hence markets for development credits could empower organizations with strong local knowledge and non-traditional approaches to problems. As long as the verification system underpinning the market works properly, the results-based approach would maintain the right incentives, achieving an expansion of local participation while ensuring high quality.

The market-based approach also helps to decentralize the funding side. In the current environment, donors need the capability to monitor programs in the field in order to be confident that their money is used well and for the uses originally intended. If small donors and charitable individuals in rich countries can purchase credits generated by verified social and environmental impacts, they

can participate with confidence in creating social value despite their small scale and limited ability to directly oversee the projects they fund.

3.4 Summary and remarks

The basic idea is simple: if an organization can present credible evidence of the social and environmental impacts of its programs against a fixed set of goals, then it gets credits corresponding to the magnitude of those impacts. If a donor wants to use its funds to generate social and environmental impacts in the developing world, it can buy credits. The impersonality and impact-orientation of the market generates better information and incentives for both implementing organizations and donors: the former are rewarded for their success, and the latter know what they are getting for their money. As a result, opportunities for generating high social returns at low cost are sought out, innovation is rewarded, and resources flow to programs which improve poor people's lives.

4 Making It Work

Our vision is for a decentralized market setting in which donors can purchase credits generated by programs and organizations that make credibly and independently verified progress on a set of core development goals. This idea is straightforward enough, but several major obstacles prevent its easy implementation. If there were not significant market failures in the way, the market we describe would already have spontaneously emerged in response to the developed world's willingness to pay for progress on human development. The most obvious of these failures are generated by the lack of credible information about the impacts of social and environmental programs and the lack of an agreed-upon system of property rights over those impacts. Like many other markets, this one requires strong and credible institutions to allow it to function efficiently if at all.

As we see it, a crediting system for verified development impacts needs at least six core institutional elements. The first is standardized set of basic outcome measures that define the various market segments. The second is a mechanism for pricing credits generated by impacts in each market segment according to the supply of impacts generated by existing programs and demand for impacts from donors and charitably-minded individuals. The third is a system for generating credible and independent estimates of program impacts. The fourth is a set of rules for

translating impact estimates into development credits, ideally in a way that accounts for statistical uncertainty arising from sampling error, potential extrapolation from outputs to outcomes, and so forth. The fifth is a definition of property rights, i.e. who owns a project or program's impacts (carbon credits under CDM take care of this through project registration rules). Together, these first five elements provide the basic information and property rights system that allows a market in verified development impacts to function. The sixth element is a set of safeguards and arbitration mechanisms to guard against potential unintended negative consequences of the incentives-based approach. These institutions might come in the form of a set of rules and policies set out by a donor operating its own crediting program, or in the form of a larger agency set up to act as a market operator for a global market for development credits akin to the carbon marketplace.

4.1 Outcomes and market segments

Metrics used in monitoring and evaluation generally consist of *inputs*, *outputs* and *outcomes*. Final outcomes are the results that we really care about; clear examples derived from the MDGs include poverty and living standards, school completion rates, child and maternal mortality, deaths from HIV/AIDS, and gender differentials on these quality of life indicators, as well as rates of deforestation, resource depletion and carbon emissions that will reduce the quality of life of future generations. Outputs are the intermediate results intended to lead to better outcomes: for instance, building schools and training teachers will hopefully result in higher rates of school enrollment and completion and higher test scores. We care about outputs to the extent that they affect outcomes, but they are not the end goal in of themselves. Inputs are financial resources, policies, procedures and methods used to generate outputs and improve outcomes. Similarly, we care about inputs to the extent that they eventually produce improvements in outcomes.

A market for development credits could be made up of several market segments, each defined by a basic outcome measure and an allocation of credits to programs that create verified progress on that measure. It is not necessary or desirable to require every program applying for development credits to verify impacts on basic outcomes directly; this point is addressed extensively below. However, basic outcomes are the right building blocks for a market for development credits. Outcomes are what we actually care about, and are simple, few in number and relatively easy to quantify. A focus on outcomes makes a wide variety of projects directly comparable: for instance, a malaria prevention project and a sanitation project both reduce child mortality and disease burdens and can be judged accordingly.

The market's benchmark outcomes should be relatively few in number and reflect the core objectives of the development community as set out in the MDGs. Good candidates include:

- The incidence of poverty, as measured by headcount, poverty gap, etc.
- maternal and child mortality rates
- basic health status as measured by disability-adjusted life years (DALYs)
- rates of literacy and school completion
- gender differentials in poverty, educational and health outcomes
- biodiversity and ecosystem productivity

The basic principle is simple. Many donors are interested in improving the living standards of poor people. Suppose in the initial market design process there was some disagreement among major donors on the best measure of poverty, but most were happy with at least one of three measures: say a headcount ratio, a poverty gap, and a convex poverty gap.²¹ The development credit market could then include three segments, and individual donors could allocate their resources to the segment that they valued most relative to the prevailing price. The market operator would introduce similar markets for health, educational and environmental outcomes according to its analysis of the donor demand for progress on those outcomes. As a result, a foundation might choose between buying credits for DALYs generated by health programs that reduce disease burdens in developing countries for \$50 per DALY and buying credits for one-dollar improvements in the annual consumption level of poor people for \$5 each. New market segments could emerge in response to donor demand, for instance as new environmental challenges arise that are not adequately covered by the existing set of outcome measures.

4.2 Verification of program impacts

Once the set of key outcomes (and hence the scope of the market) is defined, the next requirement is for a system that can provide credible information on the true impacts of development programs on those agreed-upon outcomes.

²¹ A headcount ratio measures the proportion of people below some poverty line, and hence improvements in living standards are measured to the extent that they raise some people above the poverty line. In an ordinary poverty gap measure, the weight given to an individual poor person is proportional to her distance to the poverty line, so that an incremental increase in the consumption of anyone below the poverty line contributes equally to a reduction in the poverty gap measure. In a convex poverty gap measure, increases in the consumption of the poorest people are weighted more heavily than increases in the consumption of people who are closer to the poverty line.

When consumers buy a loaf of bread, they know what they are getting with a high degree of certainty, can verify its characteristics upon consumption, and can react to inferior quality by returning the loaf or switching to a different bread vendor. If a donor or a concerned individual wishes to pay to reduce the number of Eritrean children dying of malaria over the next five years, they can contribute to one of several well-known organizations with compatible mission statements, but they can obtain very little credible, quantifiable information about the incremental impact their dollars achieve, without organizing and funding an impact evaluation themselves. Fixing this problem is crucial to allowing a market-based approach to function well.

One of the most important advances in development research and practice in the last ten years is the use of rigorous evaluation techniques to scientifically estimate the causal effects of development projects on their target populations. Medical researchers have long used controlled trials to measure the effects of drugs and clinical treatments on patient outcomes. The adaptation of related techniques in the social sciences accelerated rapidly in the 1990s, pioneered by economists like those at MIT's Poverty Action Lab and exemplified in the well-known *Progresas-Oportunidades* conditional cash transfer program in Mexico. Such studies have now provided powerful evidence on the impact and cost-effectiveness of a small number of programs and technologies introduced in the developing world. See Cracknell (2000) and Duflo (2004) for reviews of relevant methods and issues.

If it was common practice for every large development program to contain an independent evaluation component meeting scientific standards, building a market for verified social and environmental impacts would be a relatively straightforward task. However, still only a small minority of programs are properly evaluated, including large programs for which a credible evaluation would account for at most a few percent of overall costs. This is in part because of the newness of the use of these techniques in social sciences and the moderate technical capacity needed to apply them. However, it is also in part because rigorous evaluations are public goods which produce knowledge that benefits researchers and practitioners worldwide. Such services are chronically under-provided in the absence of policy mechanisms to actively promote them. We argue that an institutional framework for a market in development credits must be anchored by a system for generating information about the impacts of development programs using the tools of rigorous evaluation.

The next sub-section offers some background on rigorous evaluation, and can be skipped over by readers with significant exposure to these tools. We suggest that readers who wish to obtain more

background on rigorous evaluation read the excellent survey papers in the references below. See also the resources at www.evidencebasedpolicy.org. The following sub-section discusses how rigorous evaluation techniques could be used to facilitate a market in verified development impacts.

Rigorous evaluation

Traditional project monitoring and evaluation systems provide valuable, if not necessarily causal, information. A project designed to reduce carbon emissions and respiratory illnesses by providing energy-efficient stoves to rural households might collect information on (i) *inputs* like the amount of money spent, the number of districts worked in and the procurement practices used; (ii) *outputs* like the number of stoves distributed and the resulting reductions in indoor air pollution, and (iii) *outcomes* like fuel consumption, money and time spent acquiring fuel, and the incidence of respiratory illness. The latter are the results that we really care about and that tie closely into the MDGs.

In order to measure the causal effect of a project on the outcomes that we care about, we need a credible method of estimating the counterfactual: what outcomes for the target population would have been in the absence of the project. For instance, in evaluating the carbon emissions and health impacts of an energy-efficient stove project in a developing country, we want to know how healthy households would have been and how much fuel they would have burned had they not received the new stoves. Then if we know how healthy the households actually are and how much fuel they burn, we have a good idea of the true impact generated by the project.

Estimating an accurate counterfactual generally requires a credible comparison group. Suppose for the purpose of illustration that the project mentioned above distributed one thousand energy-efficient stoves in rural Kenya through local village heads, and that an evaluation compared the health of the households who ended up with stoves to the health of other households which did not. However, village heads might give the stoves to wealthier, better-connected households in the village. One might imagine that these households are healthier and burn more fuel because of their wealth. Then the evaluation would over-estimate the health impacts of the stoves, because the people receiving the stoves (the 'treatment group') were healthier to begin with, and would underestimate carbon reductions, because the treatment group has a higher fuel demand.

In some cases, finding a credible comparison group and estimating valid counterfactuals requires a randomized controlled trial, like commonly used in medical research. In other cases we can

construct reasonable counterfactuals from randomizing the order of a program's rollout across geographic areas, or from randomly allocating discount coupons for new technologies or for participation in fee-based programs. In some cases we can credibly identify program impacts by combining information on a sample of households or villages over a sufficiently long period of time; for instance, if a household's fuel consumption is relatively constant over several years, and then drops sharply immediately after the installation of an energy-efficient stove, while its neighbors' fuel consumption remains steady, we can be fairly confident of program impacts without randomization. In practice, economists and evaluation specialists work closely with implementing agencies to design evaluations that are practical, locally appropriate, and meet the necessary scientific standards. The crucial point is the need to incorporate the evaluation into the design of the intervention itself, with early and sustained collaboration between trained evaluation specialists and technology and project management experts.

In addition to telling us which programs are most effective, rigorous evaluations also help us learn how to make a given program work better. By varying elements of program design and implementation and studying the resulting outcomes, we can learn which variations produced the best results: which marketing strategies are most effective, how best to motivate front-line staff, what the right incentive for attendance of training sessions is, and so forth. In answering questions like these, rigorous evaluations can help programs grow more rapidly and more efficiently, attract more resources, and multiply the influence of a program and its funders by encouraging others around the world to learn from its results.

Using outputs to estimate outcomes

We mention above the distinction between inputs, outputs and outcomes. Outcomes are the best focus for measuring and rewarding the results of development programs, for several reasons. First, and most obvious, outcomes are what we care about in the end. Second, outcomes are simple, few in number and easier to quantify and judge than the many processes, procedures and policies that make up a project. Third, a focus on outcomes makes the effects of a wide variety of projects directly comparable: a malaria prevention project and a sanitation project should both reduce child mortality and can be judged accordingly. Finally, a focus on outcomes sharply reduces the potential for gaming. One often hears the concern that too much emphasis on measurement and indicators may lead to perverse incentives; for instance, an NGO paid strictly according to the number of sick patients it treats may invest in the size of its clinics at the expense of the quality of care. This is irrelevant if development credits are assigned for outcomes like reduced maternal or child

mortality: service providers are incentivized to do whatever is most effective and efficient to save lives.

However, it is not necessarily feasible to base every evaluation on final outcomes. Some outcomes we care about adjust slowly, requiring long follow-up periods with multiple surveys. This is particularly the case with environmental outcomes like ecological diversity and ecosystem productivity. In other cases outputs are under direct administrative control, and hence the counterfactual for outputs is directly known (e.g. no new beds in the clinic). If a program's outputs can be credibly and independently verified, existing knowledge about output-outcome relationships can be used to estimate the program's impact on final outcomes. The extra statistical uncertainty created by this extrapolation can be factored into the number of credits generated.

To give an example, suppose that rigorous evaluations of two projects in rural Kenya and Tanzania find that a particular insecticide-treated bed net installed in a child's sleeping area reduces the annual probability of child mortality by one percentage point on average, with some amount of variation both between Kenya and Tanzania and within different parts of each country. Suppose that another program provides similar bed nets in rural Uganda, and an independent evaluation verifies the number of bed nets properly installed. We can estimate the number of lives saved by the Ugandan program using the bed net - mortality relationship from the Kenyan and Tanzanian evaluations, with our level of confidence in the new estimates adjusted in a standardized way.²²

This puts a real premium on the body of global knowledge available from existing rigorous evaluations. As we learn more and more about the average relationship between particular sets of outputs and outcomes under different conditions, we can improve our ability to estimate the probable impacts of a program on the final outcomes we care about from what we know about its outputs. In a world where organizations are paid according to their results and the confidence with which those results are known, innovative new programs with unusual approaches have incentives to fund full-blown rigorous evaluations to fill in gaps in global knowledge, while tried-and-true programs scaling up or expanding to new areas can streamline their evaluations and focus on outputs.

²² This is not trivial. For instance, one might assume that for sufficiently similar settings (e.g. rural East African villages) the impacts of bed nets come from the same distribution up to some constant difference in means across countries, allowing one to use the dispersion in estimated impacts in different sub-sets of treatment populations to form an estimate of the range in which impacts in a third country should fall.

It is worth noting that not every project can or should be rigorously evaluated. Full-blown rigorous impact evaluations require financial resources for specialized human capital, surveys and analysis, and hence are often too expensive to be worthwhile for very small projects. They also rely on statistical tools which are not very informative with small quantities of data, so there is also a technical floor on the required project size. The minimum scale at which a rigorous evaluation is economical and informative varies considerably based on the particular case, but in many cases can be relatively low. For instance, the present authors are involved with the evaluation of a solar cooking program that plans to distribute 1,000 solar ovens in West Africa in partnership with local communities, aiming to reduce biomass fuel consumption and improve indoor air quality and respiratory health, and costing less than \$100,000. Our calculations suggest that the evaluation will have adequate statistical power, despite the small scale and low cost of the program. Even in less naturally favorable settings, creative evaluation designs and reliance on well-known output-outcome relationships can allow other small programs to be meaningfully and credibly evaluated. For the large projects financed by aid agencies and governments, the cost of a full-blown rigorous impact evaluation can be 1-2 percent of the overall project cost, a small price to pay for knowing how effective they are and how they might be improved.

Institutionalizing rigorous evaluation in development credit markets

The basic idea is that implementing agents intending to participate for in the market for development credits must build an independent evaluation into each program, and the results of the evaluation determine how many credits of what type are generated by the program. Each program aiming at the development credit market would target one or more of the market segments defined by basic outcomes. For instance, an anti-malaria program might aim to reduce mortality rates among children and reduce the burden of disease (measured by DALYs) among adults. These outcomes, or outputs which predict them reliability (in this case, the density of malarial mosquitoes in the local environment and the number of insecticide-treated bed nets installed in child and adult sleeping areas), would be built into the data collection for the evaluation.

The key thing is to ensure the credibility and independence of evaluations and reduce or eliminate the possibility for gaming. One can imagine different institutional arrangements designed to achieve this end; that outlined here is an example.

An Evaluation Committee (EC), made up of top academic experts on rigorous evaluation supported by a staff of evaluation and sector specialists, might play a central role in ensuring quality control and independence in program evaluations used to apply for development credits. The EC would

work with professional organizations of economists and evaluations experts to maintain a set of best-practice guidelines for rigorous evaluations. These professional organizations, much like those for doctors, nurses or lawyers, would develop exams for the licensing and certification of third-party evaluation consultants who would be eligible to serve as independent evaluation advisors for implementing agents intending to participate in the market. They would maintain a public database of licensees and their experience and fields of expertise.

NGOs and governments would be encouraged to build their own internal capacity to design and implement rigorous evaluations, but the credibility provided by independent third-party agents is crucial to the functioning of the market for verified development impacts. Implementing agents would select appropriate licensed advisors to help them design and implement their evaluations in compliance with the quality standards set out by the EC and the professional organizations. The licensed advisors themselves would be subject to periodic audits authorized by the EC and the professional organizations, evaluating the quality of their evaluation designs and implementation. For instance, for a program distributing insecticide-treated bed nets to fight malaria in rural areas, a team from the EC could arrange for a repeat follow-up survey for a random, comparable subset of the target population, measuring the same outputs and outcomes as the original evaluation. If the results of the repeat survey are statistically similar to the results of the original evaluation, the audit is completed. If significant discrepancies are found, a larger audit would be initiated. If after extensive inquiry the initial evaluation is found to have systematically over-inflated program impacts and no reasonable explanation is found, the implementing agent and the advisors would face legal action. The basic idea is that close monitoring of an evaluation's procedures and operations is costly, intrusive and hard to manage, so it is better for implementing agents to face some probability of an ex post audit that will uncover any significant fraud directly.

4.3 Translating impact evaluations into development credits

When an evaluation for a completed program is certified by the Evaluation Committee and submitted to the market operator, the latter must determine how many credits of different types the program generates. The primary issue is the level of statistical confidence of the evaluation in question. An estimate of a program's impact on a particular outcome comes in two parts: a point estimate (or best guess of the impact) and a standard error (or level of uncertainty of the impact). When comparing two programs with the same point estimates for impacts on living standards, it seems intuitive that we prefer the one with lower standard errors around that estimate. Building this value of precision into the formula by which development credits are allocated incentivizes

implementing agents to design high-quality evaluations and ensures that donors feel confident in what their money is buying.

There are two main things which introduce uncertainty into impact estimates. The first is sampling error. Rigorous evaluations use data collected on the target populations to estimate how the outcomes of interest change upon receiving the intervention. Many changing factors affect the social and environmental outcomes we care about, and for an evaluation with a finite sample size we can never perfectly pin down the exact impact of the program. However, we can measure the uncertainty generated by sampling error, and trained specialists can forecast it in advance for a specific program and evaluation design, so implementing agents can factor it into their financial analysis and planning.

The second factor introducing uncertainty into evaluation results is extrapolation from outputs to outcomes. Programs which can convincingly measure effects on outcomes need not worry about this. For others this may be infeasible or uneconomical, and they may instead provide evidence of outputs, like the number of new math textbooks delivered to a rural school district or the number of insecticide-treated bed nets installed in child sleeping areas. If the existing body of knowledge from rigorous evaluations includes estimates of the relationship between new bed nets installed and child deaths from malaria averted from a similar malarial region as the candidate program, that average relationship can be used to estimate the impact of the candidate program, and the standard errors of the final impact estimate adjusted upward accordingly.²³ The result would effectively be to set “exchange rates” between specific outputs and the basic outcomes that generate credits, determined by the confidence we have in the output-outcome relationship in question. This incentivizes agents to invest in rigorous evaluations for innovative new approaches and to rely on outputs when the output-outcome relationship is well documented and measuring outcomes directly is difficult.

These two sources of uncertainty combined generate a composite standard error and confidence interval for any impact estimate. There are several reasonable ways to ‘price’ this residual uncertainty. In the academic world it is common to treat a point estimate that is significantly different from zero with 95% probability as adequate evidence of a positive effect. However, too-rigid adherence to traditional notions of statistical significance is unwise. The probabilistic

²³ This is not as simple as it sounds. The main adjustment comes from the standard errors associated with the output-outcome relationship. However, an additional adjustment should ideally be made for the fact that it is a different program in a different place which may have a different output-outcome relationship. There is no consensus on the best way to approach this, and the methods available depend on the assumptions one is willing to make.

difference between T-statistics of 1.8 and 2.0 is very small, and discontinuities in the credit system induced by a threshold-based approach would be unreasonable and would encourage cheating. A better approach would be to assign credits based on the left end of some confidence interval (say 90%, which is equivalent to the point estimate less about 1.65 standard errors). This would be a conservative approach that generates strong marginal incentives for implementing agencies to design evaluations with strong statistical precision. If there was too much disagreement about the value of uncertainty among donors, one could construct separate high-risk and low-risk market segments with different formula for credit allocation in each.

4.4 Pricing mechanisms

Once the key outcomes are defined and a system is in place for translating evidence of program impacts into tradable credits, the market needs a way to price those credits. A pricing system needs to be able to clear the market as efficiently as possible, ensuring that there are enough buyers for the credits offered at any given time, and to minimize price volatility and uncertainty to market participants. Pricing mechanisms which are more voluntary and decentralized in nature are attractive because they remind us of the markets for consumer goods that we are used to, but they would probably not perform well on these criteria.

Voluntary matching. In spot markets for consumer products, buyers compare the features and prices of products offered by suppliers and choose which to purchase according to their preferences, either negotiating or taking the offered price as given. This could be emulated by a voluntary matching system in which donors identify organizations they like and negotiate a price for the credits which their programs have generated. For instance, an NGO wrapping up an anti-malaria program in rural Tanzania would post its independently evaluated results to a centralized database, and a USAID portfolio manager responsible for maximizing the development impact of a \$100m fund with respect to the priorities of the US government could see the program and decide to negotiate with the NGO to purchase the credits generated by its impacts on DALYs and child mortality.

This type of approach would be easy to implement, requiring only a set of evaluation standards, a website and a large marketing and publicity campaign, like a more detailed GiveWell.net with a rigorous evaluations focus. However, while this might be a valuable incremental step, we argue that it cannot achieve the central goals of the present proposal: to create a world in which on-the-ground results are directly verified and rewarded in an impersonal fashion disconnected from

current fads or trends in the donor community, agents' presentation skills and ability to build relationships with donors, and so on. This approach retains much of the subjectivity and uncertainty of the current funding environment, as implementing organizations must depend on finding one or more donors who want to specifically fund their work. Market power would be a real problem because only a handful of large donors at any given time would be interested in a particular type of program in a particular country.

Supplier pricing. This is similar to voluntary matching, but implementing organizations would post take-it-or-leave-it prices for credits generated by their programs. The absence of negotiations reduces incentives for strategic behavior, but donor market power would still be a problem, as donors choose individual organizations to fund rather than impersonally financing verified results regardless of where or how they were generated.

Donor pricing. Donors could post take-it-or-leave-it prices at which they are willing to purchase given quantities of different types of credits, and implementing organizations could choose the donors which offer the highest prices for their credits. This is a substantial improvement for implementing organizations because their credits are valued independently and uniformly at the market price, as in the case of carbon credits. Donors are forced to compete with each other in a few large market segments, ensuring that any one donor cannot dominate the price-setting process.

Bid pricing. A more centralized mechanism might borrow from the example of wholesale electricity markets.²⁴ The operator would set a market interval, perhaps monthly or even quarterly so that a sufficient number of new programs complete their evaluations and are assigned credits for each market segment during each interval. At the end of each interval, each implementing agency would submit a bid sheet listing the minimum price it will accept for its credits and how many credits it is willing to sell for a range of market prices. Each donor would submit a bid sheet listing the quantity of credits of each type it is willing to buy for a range of market prices. These bids could be complex if desired, but in practice would often be relatively simple (e.g. "buy credits equivalent to 20,000 DALYs for a price of up to \$30 per DALY, and 10,000 DALYs for a price of up to \$50"). The

²⁴ In these markets, the operator sets a market interval, in most cases a half hour. Suppliers submit bid sheets, which commit them to sell some quantity of electricity on the market as a function of the market price, each day for the next day's forty-eight half-hour periods. For instance, a simple bid for a specified half-hour period might commit zero megawatts for a price below \$0.03 per megawatt, up to five megawatts for a price between \$0.03 and \$0.05, and up to seven megawatts for a price above \$0.05. The market operator takes all the bids for the day, and then calls in all the generator capacity necessary to meet demand during every half-hour period, starting with the lowest-bidding generators first. The price paid to all firms providing power during that half-hour period is the price required to call in enough capacity to meet demand given the bids submitted the previous day.

price in each market segment for that period is the price which sets supply equal to demand. Competition should be adequate as long as sufficient numbers of large donors participate in the market.

One might reasonably raise the concern that the prices paid for results in different areas are set by donor demand and hence may be influenced by fads and politics.²⁵ This is certainly true to some extent. However, market forces push back against price distortions induced by the fickle winds of donor preferences. For instance, suppose that for some reason many major donor agencies shift their resources into health and education markets and out of income generation and poverty reduction. Then the price of health and education impacts will rise and the price of living standards impacts will fall, leading other donors to shift resources into buying credits in the living standards markets upon seeing the great value available.

One might also wonder whether it would be better to segment the market by country, allowing donors to specifically purchase credits generated by poverty reduction programs in Kenya. This is an intermediate option between the purely voluntary matching market and our more institutionalized vision. In one sense, the greater choice available in such a setting is a good thing, as a donor who prioritizes progress in a certain region is able to allocate its funds accordingly. However, this also brings several drawbacks. In terms of philosophy and political feasibility, the idea of putting an explicit market price on averting a child's death is likely to be shocking enough to the average citizen. Allowing different prices for the lives of children in Kenya and China according to donor priorities in a given year would be hard to explain to stakeholders and politically problematic.²⁶ Second, the resulting markets would be much sparser. This raises the risk of market power, as one large donor with a strong interest in a particular set of countries could easily dominate their development credit markets. It also imposes more risk on implementing agencies: a program operating in a small country would have few potential buyers, so NGOs would tend to emphasize larger countries where demand is more predictable.

²⁵ This raises a crucial point about a market for development credits: the demand side arises from the preferences of wealthy people over improving the lives of others and from the political value to elected officials of putting resources into development programs. Hence this demand curve has different welfare properties than a standard market demand curve for consumer products. The latter corresponds directly to welfare because it derives from the utility-maximizing choices of consumers: someone who is willing to pay \$100 for a bicycle and pays a market price of \$75 receives net welfare (consumer surplus) of \$25. In contrast, we generally think of poverty reduction as a common social goal with some (possibly implicit) unit value at a point in time, rather than as a consumption good for wealthy people.

²⁶ On the other hand, this already occurs within the current aid system where aid will spend more to save a life in richer or high priority countries compared to how much they are willing to spend in difficult areas, poorer countries or lower priority regions.

4.5 Perverse incentives and unintended consequences

In this section, we mention various adverse or unintended consequences that may result from an incentive-based development credit market system like that described in this paper. While efficient incentives may assist greatly in allocating resources in ways that are well-aligned with resource-allocation goals, there is also the potential for gaming. We discuss a few of the potentially most severe types of games and unintended consequences that might occur in a development credit market system. These may include:

1. Manipulating the baseline
2. Jumping in and out of the market
3. Selective reporting and monitoring
4. Rent-seeking behaviors
5. Restricted access to capital
6. Free riders and non-additionality

We discuss each of these market games and the potential consequences in the following subsections.

Manipulating the baseline

Since impacts are measured relative to some baseline, manipulation of the baseline could potentially increase a program's measured impacts and hence the financial flows to its implementers. In a hypothetical but potentially horrendous example, a government or project implementer could intentionally cut budgets for health services in some locality, allowing health outcomes to deteriorate before the baseline survey so that its subsequent programs might earn more credits. A potential counter-measure for this type of gaming is to require adequate documentation of program budgets and program activities for one or more years prior to the baseline survey. The market operator could also set historical development impact liability as a requirement for market participation of any entity receiving revenues from development credits. Such a liability might set the baseline with an anti-backsliding provision that says that the baseline is the best historical development condition since the beginning of the market. Some provision for mitigating at least a portion of historical backsliding could be charged from future credit revenues. The existence of such a historical liability can potentially provide incentives for good behavior for actors that are thinking about future participation in development credit markets.

Jumping in and out of the market

Another form of gaming in a development credit market can come from jumping in and out of the market depending on the degree of success of different projects delivering development impacts. Naturally, market actors will have the most motivation to participate in a market when their projects are most successful. Yet other actions that they may take outside of the market may more than undo the social, environmental, and development benefits that they produce from projects that they put into the market.

This type of behavior is already seen in U.S. carbon voluntary offset markets. There are companies which at the same time build multi-billion dollar coal-fired power plants which can contribute dramatically to global warming and climate change, yet at the same time the same company may sell carbon credits from wind power projects to help consumers voluntarily “offset” their climate emissions.²⁷ The company is not taxed for its increased emissions in one area, while receiving incentives for progress in another area. Without a carbon tax on the activities that contribute to climate change, the company can earn a net increase in profits from making the environmental problem worse rather than better.

The analogous project with health credits might be a pharmaceutical company that on one side of the ledger denies distribution of essential drugs to developing country markets because of the potential impact on rich-country sales, yet which earns health credits for the distribution of low-priced generic drugs for particular diseases in developing countries. Similarly, without a tax on the activities that prevent disease reduction, the company may earn greater net profits by restricting distribution of essential drugs until it can get some subsidy from health “credits.” In this case a credit market would increase profits and provide incentives for companies with socially bad behavior.

A possible counter-measure to this type of gaming is to set appropriate requirements for market participation that any entity receiving payments for credits must satisfy. Such requirements can

²⁷ For example as of April 2008, the company America Electric Power (AEP) was simultaneously developing large coal-fired power plant projects while selling CO₂ emissions offsets for its clean energy projects through www.carbonfund.org. Allegedly consumers could buy these credits as non-profit donations to offset their greenhouse gas emissions but their donations were being paid to a company which in net terms is making the climate change problem worse.

give a competitive advantage to project developers that are focused on producing positive development impacts in all of their activities.

Selective reporting, monitoring, and information access

Project developers and implementers in a development credit market may have an incentive to erect selective barriers to information access for their worst performing projects. The direct effects of this are small, as projects with insufficient evidence of impacts will not receive credits. However, selective information may produce biased information on the relationships between inputs and outputs which would otherwise be valuable for the operation of the crediting system as a whole. Through information, reporting, and auditing standards and project pre-registration at a central repository the market can provide the means for studying, evaluating and correcting market inefficiencies brought by project developer reporting and evaluation biases. In addition, this suggests that the market system might operate a fund in conjunction with governments and international institutions which finances rigorous impact evaluations by independent academics focused on key areas where knowledge is deemed insufficient. Such studies are a valuable public good for the market as a whole, and have the potential of addressing potentially very large market inefficiencies.

Rent-seeking behaviors

Rent-seeking behaviors can arise when a system allows unaccountable or unmonitored individuals a key role in the approval and distribution of project benefits. A social and environmental crediting market is likely to generate whole new sets of requirements and rules for project registration, monitoring, evaluation and crediting. And with each new procedure and regulation, the enforcer, approver, or evaluator of that procedure or regulation becomes a potential rent-seeker that can charge a fee or make unrelated demands for approval or consent. Hence the creation of a development credit market could conceivably lead to new forms of corruption in weak and failed states.

It might also increase the ability of rich country technical consultants to charge high fees (a legal form of rent) to help implementing organizations navigate the procedures and regulations. In carbon credit markets, the technical requirements for project registration and certification have lead to few projects being implemented in poorer regions such as sub-Saharan Africa. A key reason for this is the extremely high transaction costs of project registration, monitoring and verification,

and the fact that most experts and consultants in carbon credit markets are from the richer countries buying the credits.

These issues put a large premium on streamlined regulations and procedures which ensure easy-access, low-cost competition. They also suggest the importance of high-quality training programs for developing local monitoring and evaluation capacity in poor countries. The existence of the market should create incentives for public and private actors to invest in such programs, but coordination through and certification by well-known international organizations would assist greatly with this.

Restricted access to capital

A simple shift from grant aid to development credits that does not address the capital financing needs of project developers also has the potential of increasing capital supply shortages for developing countries. Business, organizations and local NGOs in developing countries often have very limited access to capital at high interest rates. Because of the uncertainties and complications of revenues from credits in a new development credit market, it may be very hard for project developers to use such potential revenues as collateral for capital financing. Under such conditions, rich country investors who specialize in development credit financing will be able to demand a large share of project credit revenues in exchange for financing development project implementers. High capital costs will deter new entrants who need start-up funds and raise the risks facing smaller implementing organizations which cannot easily smooth their cash flows. This reduces the efficiency advantages of the crediting system while allowing rich countries to capture much of the remaining surplus.

The problem of weak capital markets suggests a vital role for foundations in providing cost-effective capital financing mechanisms for project development along with market access for developing country project implementers. This financing will be most effective if it allows local implementers to leverage local equity capital with low cost loans thus allowing greater initial investments in project assets.

Free riders and non-additionality

Free riders and the financial and motivational non-additionality of development credits is perhaps the hardest problem that a development credit market will need to resolve. Free riders are those who take advantages of incentives even when they would have implemented the action without the

incentive. And financial non-additionality is where a project would have been implemented anyway with existing resources with any credits earned not contributing to either the production of credit or the continuing replication of the project impacts.

The key questions that financial additionality addresses are: What if those projects that are best at creating cost-effective development impacts are those which would have produced the social benefits anyways for other, unrelated reasons? Are the development impacts actually funding development or are they simply distributing money to no effect? This is also perhaps the biggest issue with carbon credit markets. Since a controlled experiment cannot be performed on a single project to determine whether such a project would have been performed without a credit market incentive, in most cases financial and motivational project additionality for development credits cannot be scientifically determined.

The resolution of this issue rests in two different types of approach. One approach used by current carbon credit markets is an administrative one where the project developer documents the financial structure of the project and shows that the project finances and organization meets criteria and tests that certify additionality. This may be an investment return analysis or this may be in the form of a barrier removal analysis. The weakness of this approach is that it does not measure non-financial reasons for project implementation, nor does it explore the total universe of project implementation and financing options and show the incremental impact of credit financing compared to a non-credit alternative. In reality, project developers may often select a financing package for a project that is structured so that they can qualify for the extra revenues from carbon credits even though they have other feasible project financing alternatives without carbon credits.

One type of study that could be conducted in a very sophisticated development credit market that also offers credit-based financing packages for projects developers would be to measure actual project implementation elasticities as a function of credit price or financing terms. These could be done by adding a random component to the selection of set of financial aid applications from project developers and the terms of financial aid. This would then provide a financing treatment and control group that could be evaluated with respect to actual project impacts. Then for those donors who wish to fund only strictly additional development impacts, the impact per output relationship could be adjusted to reflect financial additionality effects.

This then could provide a rigorous evaluation method for evaluating the impact of impact-based credit financing. This more rigorous method would be the ideal resolution for the free rider and the financial additionality issue.

5 Summary and Conclusion

This paper describes the notion of a market that directly rewards verified social and environmental impacts in developing countries. Such a market would operate much in the same way that carbon credits are generated by verified reductions in carbon emissions.

A development credits system would bring important advantages, including shifts in the allocation of development finance towards more cost-effective programs and organizations, stronger incentives for innovation and for good management, less centralization, reduced transactions costs, and ultimately more impact for each dollar spent. However, such a system, especially in its most ambitious form, requires a strong institutional framework capable of providing credible and transparent information about program impacts, efficient prices for impacts, effective capital financing mechanisms fraud deterrence and mitigation of unintended negative consequences and perverse incentives. This paper has examined these advantages and challenges in the hope of generating a constructive policy dialogue.

The conclusion that a development credit market can produce very large benefits rests on a set of assumptions and hypotheses about how such a system might function and the impact such a market might have on donors and project implementers. These assumptions include:

1. A substantial number of donors and individuals will want to participate in such a development credit market in order to increase the effectiveness, transparency, and efficiency of their donor contributions.
2. A diversity of existing and new project implementers will wish to participate in a development credit market because either it is easier for them to prove impacts than go through traditional grant financing, or because they find that their projects are more cost effective than average and therefore more 'profitable' in a development credit market.
3. Rigorous evaluation can tie impacts to project outputs and through tracking project outputs, and project implementers will be able to estimate project impacts that can be credited.
4. Informal equity capital, or effective forms of loan financing will be available for small, yet efficient actors that wish to enter into a development credit market.
5. Market gaming and unintended consequences can be kept under control through market monitoring and the right institutional evaluation and regulatory frameworks.

6. Prices and project impacts can be stable and predictable enough to provide continuing incentives for project implementers to continue measuring and funding projects based on development credits.
7. Issues of co-financing, free riders and financial additionality can be resolved.

These assumptions effectively define a set of conditions under which a development credit market can develop. For there to be a development credits system both donors and project implementers will have to participate; rigorous evaluations will be needed to tie observable project outputs to development impacts; capital will need to be available to project implementers in a way that makes project development attractive; market gaming and unintended consequences will need to be kept under control; prices will need to be sufficiently high and stable to provide a continuing incentive for market participation for implementers; and the market will need to provide a credible financial impact on projects that allows them to grow and replicate at rates that are higher than what non-participation would entail.

The authors believe that the above conditions for the creation of a development credit market can be met, and hope that discussions arising out of this paper can help refine these issues and requirements and advance progress toward impacts-based funding in development. But ultimately the true test of the development credit market concept is through a pilot test and demonstration. The investigation and evaluation of a test-market will provide much clearer and more certain answer regarding possible market functioning and efficiency.

In a follow-up paper the authors hope to describe and discuss such a development credit market pilot test. By linking rigorous evaluation of development project impacts to efficient financing mechanisms based on those impacts, the authors believe that institutions can be built that can dramatically increase the efficiency and effectiveness of development aid in general.

6 References

Akin, William E. (1977). *Technocracy and the American Dream: The Technocrat Movement, 1900-1941*. University of California Press.

Arvin, B. Mak (2002), *New Perspectives on Foreign Aid and Economic Development*, Praeger Publishers.

Barnes, Douglas; Openshaw, Keith; Smith, Kirk R; van der Plas, Robert (1994), *What Makes People Cook with Biomass Stoves?*, Technical Paper no 242, Energy Series, Worldbank, 1994, Washington, DC.

Capoor K. & Ambrosi P. 2007. "State and trends of the carbon market 2007." The World Bank and IETA.

Collier, Paul (2007), *The Bottom Billion: Why the Poorest Countries are Failing and What Can Be Done About It*. Oxford University Press.

Collier, Paul and Gunning, Jan (1999), "Explaining African Economic Performance," *Journal of Economic Literature*, 37, 1: 64-111.

Cracknell, Basil Edward (2000), *Evaluating Development Aid: Problems, Issues and Solutions*, SAGE Publications

Dichter, Thomas (2003), *Despite Good Intentions: Why Development Assistance to the Third World Has Failed*. University of Massachusetts Press.

Dollar, David and Levin, Victoria (2006), "The Increasing Selectivity of Foreign Aid, 1984-2003," *World Development* 34, 12: pg. 2034-2046

Duflo, Esther; Glennerster, Rachel; Kremer, Michael (2006), "Using Randomization in Development Economics Research: A Toolkit," CEPR working paper #6059

Duflo, Esther; Kremer, Michael (2004), "Use of Randomization in the Evaluation of Development Effectiveness," in *Evaluating Development Effectiveness*, Feinstien, Ingram & Pitman (eds), vol. 7, p. 205-232, New Brunswick, New Jersey & London: Transaction Publishers

Easterly, William (2003), "Can Foreign Aid Buy Growth?" *Journal of Economic Perspectives*, 17, 3: 23-48

Feinstein, Osvaldo; Ingram, Gregory; Pitman, George (eds) (2004), *Evaluating Development Effectiveness*, The World Bank Series on Evaluation and Development, Vol. 7

Hamilton K, Bayon R, Turner G & Higgins D, "State of the voluntary carbon market 2007 – picking up steam," *Ecosystem Marketplace & New Carbon Finance*

Harris (2006), "The Voluntary Retail Carbon Market: A Review and Analysis of the Current Market and Outlook," MSc thesis at Imperial College of London.

Harrison, G. and List, J.A. (2004), "Field Experiments," *Journal of Economic Literature*, 42: 1013-1059.

Kaldor, Nicholas (1961), "Capital Accumulation and Economic Growth," in Lutz and Hague (eds.) *The Theory of Capital*, New York: St. Martin's Press

Kremer, Michael (2003), "Randomized Evaluations of Educational Programs in Developing Countries: Some Lessons," *American Economic Review* 93(2), 102-106

Miguel, Edward and Kremer, Michael (2004), "Worms: Identifying Treatment Effects in the Presence of Externalities," *Econometrica*...

Ndulu, Benno; O'Connell, Stephen; Bates, Robert; Collier, Paul; Soludo, Chukwuma (2007), *The Political Economy of Economic Growth in Africa, 1960-2000*, Cambridge University Press.

Pack, Howard and Pack, Janet Rothenberg (1993), "Foreign Aid and the Question of Fungibility," *The Review of Economics and Statistics*, 75, 2: pg. 258-265.

Rajan, Raghuram and Subramanian, Arvind (2007), "Aid and Growth: What Does the Cross-Country Evidence Really Show?" *Review of Economics and Statistics*, forthcoming.

Ravallion, Martin (2006), "Evaluating Anti-Poverty Programs", forthcoming in Handbook of Development Economics vol. 4, Robert Evenson and T. Paul Shultz (eds.), Amsterdam: North Holland.

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

Sachs, Jeffrey (2005), *The End of Poverty*. Penguin Press.

Stiglitz, Joseph (2003), *Globalization and its Discontents*, W.W. Norton

Tietenberg, Tom and Nick Johnstone (2004) "Ex-Post Evaluation of Tradable Permits: Methodological Issues and Literature Review" in Organization for Economic Co-Operation and Development. *Tradable Permits: Policy Evaluation, Design and Reform*.

Van de Walle, Dominique; Mu, Ren (2007), "Fungibility and the Flypaper Effect of Project Aid: Micro-Evidence for Vietnam," *Journal of Development Economics* 84, 2: 667-685.

Wara, Michael (2007), "Is the Global Carbon Market Working?" *Nature* 445, February 2007.

Wolfensohn, James and Bourguignon, Francois (2004), *Looking Back, Looking Ahead*. The World Bank. Available at <http://www.worldbank.org/ambc/lookingbacklookingahead.pdf>