BASELINES FOR GREENHOUSE GAS REDUCTIONS: PROBLEMS, PRECEDENTS, SOLUTIONS

prepared for the Carbon Offsets Unit, World Bank

Kenneth M. Chomitz
Development Research Group
World Bank

background materials by Hagler-Bailly Consulting, Inc., Tracy Lewis, and Environmental Law Institute

<u>Draft for discussion:</u> rev.1.4, 16 July 1998
Comments welcome: kchomitz@worldbank.org

Acknowledgements: Much of section 7 on lessons from demand-side-management is based on a background paper by Daniel Violette and Shannon Ragland of Hagler-Bailly Consulting. The discussion of the US emissions reductions credits draws on a background paper by Byron Swift of the Environmental Law Institute. The menu-choice method for baseline revelation was proposed in background material by Tracy Lewis. I am grateful to Johannes Heister and Charles Feinstein for helpful discussions and comments. However, all interpretations, conclusions, and errors are mine.

SUMMARY

Rigor in baselines

It's important to establish the right degree of rigor in baselining. Overly lax baselines will threaten the system's credibility and usefulness, and shift rents from high quality providers to low quality providers of offsets. Overly stringent baselines will discourage valid projects and drive up project costs.

The only 'magic bullet' for baselining is to set up a national or sectoral baseline, and define offsets against this baseline. A variant is to use facility-level prior output as a baseline, in a context where sectoral emissions are capped, and the caps are binding. (The US market for NOx and VOC offsets provides a precedent.) This will be difficult in most cases; in fact, joint implementation is a device for avoiding the difficulties of setting the sectoral or national caps. However, it is worth thinking about: a) in EITs; b) where project-level interventions have sector-wide implications, as in the power sector and land-use sector. In these cases, calculations of sectoral-level baselines have to be performed anyway.

Keeping baselines honest

Failing that, baseline determination unavoidably has a judgmental component. This means that baseline determination depends not just on methodology, but on a set of institutions that keep the methodology's application reasonable and honest.

Third party certification may not by itself yield unbiased results. In any situation where there are reasonable doubts, incentives will encourage practitioners to rule in favor of higher baselines. This has been true, for instance, in evaluations of public transit systems in the US, where ridership projections have consistently been biased upwards and cost projections consistently biased downwards, resulting in biases in favor of heavily-subsidized capital-intensive rail systems. In contrast, the US system of DSM incentives successfully uses panels of public interest representatives to review evaluations of net energy savings (ie. the equivalent of offset measurement) by third party evaluators. This is noteworthy because DSM incentive programs constitute a large scale (approximately $3 billion/year) analog to the carbon offsets market, facing very similar baseline problems.

Three methodological issues

The methodological issues in baseline-setting, broadly are:

1. *additionality:* the determination of which technology would have been adopted in the absence of offset sales
2. *direct emissions:* determination of direct emissions conditional on technology,
3. *leakage:* determination of indirect impacts on emissions.

Of these issues, the second is the most straightforward, though not necessarily simple. It is largely a question of measurement and sampling techniques. Detailed protocols for this exist in the energy and forestry sectors.

Additionality

Issue 1, the additionality issue, is perhaps the most difficult and subjective. There are two basic ways of making this determination:

a) using comparison groups – this may be appropriate for projects, such as DSM, which can be thought of as 'bundles' of smaller activities and for which adequate populations of control units exist.
b) simulating the project investment decision – this is an unavoidable approach for large projects without obvious comparison groups, and will probably be a feature of most PCF project evaluations. The question is 'simply': what project, if any, would the project sponsor have undertaken in the absence of offsets funding? The approach is to apply behavioral and/or financial models to predict, in a structured way, whether the proposed project would have been spontaneously undertaken in preference to the baseline or reference project. In many cases this will be equivalent to the incremental cost approach of the GEF, though with a more transparent treatment of incremental benefits.

Behavioral/financial models
The behavioral/financial modeling approach to additionality subsumes the 'barriers' approach to JI. The latter simply nominates a qualitative list of problems which raise project costs and risks. The behavioral/financial modeling approach imposes some rigor by requiring the systematic quantification of those costs and risks.

In order to apply the behavioral/financial approach we need three components:

- an *engineering or cash flow model* showing expenses and revenues under different assumptions; this could be anything from a simple spreadsheet to a complex engineering model of a facility,
- a *normative decision model* which chooses among projects based on the output of a). This decision model could be a simple spreadsheet-based comparison of NPV's or IRR's, or it could be a more sophisticated model incorporating multiple goals and constraints. At the sectoral level, a integrated resources planning model of electrical generation capacity expansion might be applicable.
- A *set of key parameters* to input into a) and b), including capital costs, expected future fuel prices, and pollution charges.

Most of these key parameters are either known only to the project sponsor, are subject to deliberate manipulation by government policy, or are subject to change over time. Default specification of these parameters, probably on a country-by-country basis, reduces the danger of moral hazard and minimizes 'gaming'. Some of these decisions, while crucial to baseline determination, cannot easily be made on empirical grounds – e.g., whether or not to accept policy-based distortions in energy prices, or how to assess future levels of enforcement effort of pollution and forestry laws. Pending any official ruling on these issues, the PCF management will have to make provisional decisions.

Partial crediting strategies
Partial crediting strategies can be used to account for uncertainty and asymmetric information in additionality or baseline determination. For instance, where the investment decision model suggests that the low-carbon project adoption is possible but marginal, partial credit could be requested. Menu-choice revelation mechanisms could possibly be applied in conjunction with the independent stakeholder review process mentioned above. These mechanisms involve letting the project sponsor choose between a high baseline with partial credit for measured reductions or a low baseline with full credit for measured reductions.

Dynamic baselines
Dynamic baselines (that is, adjustable over time) are feasible, and have been used successfully in establishing net energy savings for DSM incentives. They are desirable:

- in replacement/retrofit projects, when retirement of the existing facility is sensitive to unpredictable changes in prices or interest rates. A static baseline would require predicting a precise date when the retirement decision would have taken place in the absence of the project. A dynamic baseline is advantageous if there is a good chance that economic conditions would militate against the retirement decision for a longer-than-expected period.
- when emissions are volatile because of variable and unpredictable facility loads. A static prediction of loads would lead to greater variance in offset production – eliminating them, for instance, if loads were greater than anticipated.

The potential benefits of dynamic baselines have to be weighed against the greater costs.

Leakage
Leakages -- often discussed in connection with forestry projects -- are potentially worrisome also for fuel-switching and efficiency-increasing projects. Project-level reductions in the demand for fuels can have a 'snapback' effect as other consumers react to slightly depressed prices by slightly increasing consumption. On the other hand, positive spillover effects can amplify emissions reductions if project-sponsored technologies diffuse to nonproject facilities. General adjustment or discounting parameters for this purpose should be developed.

Duration of abatement/sequestration and forestry projects
Forestry projects have different durations of impact than do industrial emissions abatement projects. Abating a ton keeps it out of the atmosphere for the average residency time of the GHG in question. Sequestration, or deforestation prevention, is always potentially reversible. The difference needs to be explicitly accounted for when assessing baselines and calculating offsets. One solution is a "pay-as-you-sequester" scheme, in which sequestration services are reckoned on a ton-year basis (keeping a ton out of the atmosphere for a year), and credited at regular intervals. A conversion factor would relate ton-year credits to 'perpetual' tons, using a discounting formula. This facilitates setting up sequestration projects in situations where political and implementation risks discourage long term (20 or 30 year) contracts.

1. INTRODUCTION: BASELINES AND WHY THEY MATTER

AN AWKWARD BUT POTENTIALLY FEASIBLE INSTRUMENT

The Kyoto Protocol to the Framework Convention on Climate Change allows developed countries to sponsor greenhouse gas (GHG) reduction projects in host countries. The difference between the project's actual emissions, and the hypothetical emissions had the project not been implemented, constitute a savings or offset, typically measured in tons of CO2 or carbon equivalent. This quantity can be sold as an emissions reduction[1] (ER) to developed-country buyers, who use it to offset their own GHG emissions. The credit has value because the buyers face either a tax or a limit on net GHG emissions.

This device, also known as Joint Implementation (JI), is an awkward but potentially feasible solution to an otherwise intractable problem. The problem is to reduce the world social costs of GHG reduction by taking advantage of the perceived large supply of low-cost reduction options in the developing world. An international system of tradable emissions permits could accomplish this without the troublesome mechanics of defining and agreeing on hypothetical baselines. The disadvantage of such a system is the perceived current political impossibility of agreeing on emissions budgets for developing countries – in part because of the serious distributional issues involved, in part because of resistance to the concept of emissions permit trading. ER trading substitutes a large number of small and ostensibly technical determinations about project-level baselines for a small number of large, overtly political negotiations about country-level emissions budgets.

THE PROBLEM WITH BASELINES

The distinguishing feature of an ER system is that it is based on an unobservable commodity: the difference between observed GHG emissions by a project host and those which hypothetically would have occurred, had there been no project. The Protocol makes clear that reductions must be "additional to any that would otherwise occur". To define an ER, it is necessary to specify the hypothetical, unobservable baseline level of emissions.

---

[1] This term potentially encompasses both the emissions reductions units of the Protocol's article 6, and certified emissions reductions of article 12.

Agreed-on baselines will always be problematic for three reasons. First, it is inherently difficult to predict what would have happened in the 'but-for' world. Second, both buyers and sellers of ERs have strong incentives to overstate the baseline level of emissions, since this increases revenues for the seller and in aggregate may reduce the price of offsets for buyers. Third, baseline setting requires some assumptions about national policies. The project-level approach to emissions reductions obscures, but does not really eliminate, the political issues associated with setting national emissions budgets.

OVERVIEW OF THE PAPER

The next section reviews the consequences of inaccurate baselines and discusses the tradeoffs associated with different levels of baseline rigor. Section 3 focuses on how asymmetric or uncertain information about key behavioral parameters leads to baseline uncertainty. Section 4 discusses four general methodological approaches to overcoming these problems and establishing baselines. The fifth section discusses the use of partial crediting and information revelation strategies to correct for asymmetric information problems. Next there is a discussion of spatial and temporal boundary issues. Section 7 discusses the lessons learned from demand-side-management incentive programs in the US, an interesting large-scale analog to GHG offsets. The concluding section offers recommendations for baseline practitioners.

2. CONSEQUENCES OF INACCURATE BASELINES

Does it matter if baselines are overstated? Some argue that a concern with baseline accuracy reflects only a desire of developed countries to reduce transfers to developing countries. But baseline inaccuracy has wider-ranging impacts on world welfare and on income distribution among developing countries. As a benchmark, consider Figure 1. This shows the now-familiar diagram of the benefits of ER trading when baselines are accurately known. The demand curve represents the developed country marginal abatement cost curve. The supply curve represents the supplying countries' marginal abatement cost curve. The shaded area represents the abatement cost reduction realized by permitting trade in ER's. With the curves shown in the diagram, most of these gains accrue as rents to the supplying countries.

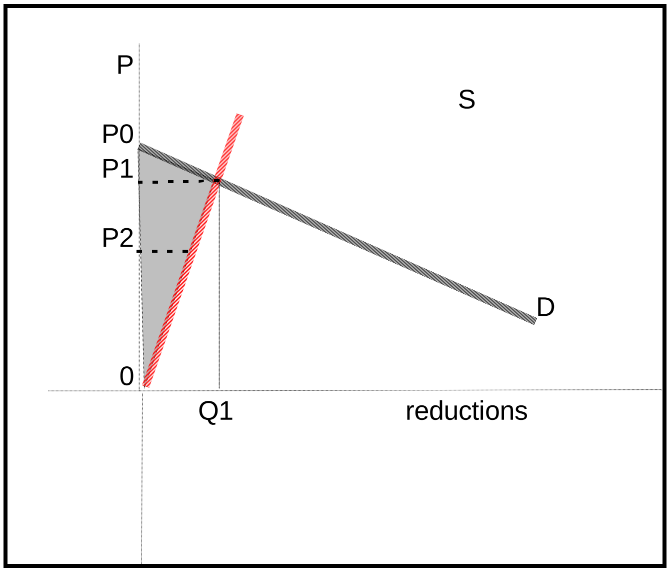

Figure 1

Figure 2 shows the outcome if some non-additional projects are presented -- and accepted -- as additional. These are likely to be economically and environmentally desirable. But because these projects, by definition, would have occurred anyway, their supply price for (non-genuine) ER's is zero. Their inclusion pushes the supply curve outward. As a result, world GHG emissions increase by Qn, because buying countries are allowed to increase their emissions by this amount. Presumably the damages from these emissions exceed the savings in buying-country abatement costs (the area under the demand curve from 0 to Qn). The gains from ER trade, shown again by the shaded triangle, are reduced. Some relatively-high cost suppliers of genuine ERs are now crowded out of the market. Remaining suppliers of genuine ERs see their rent per ton reduced by P_1-P_2. *In sum, overstated baselines result in increased GHG emissions, reduce the gains from ER trading, and divert rents away from projects with more accurate baselines.*

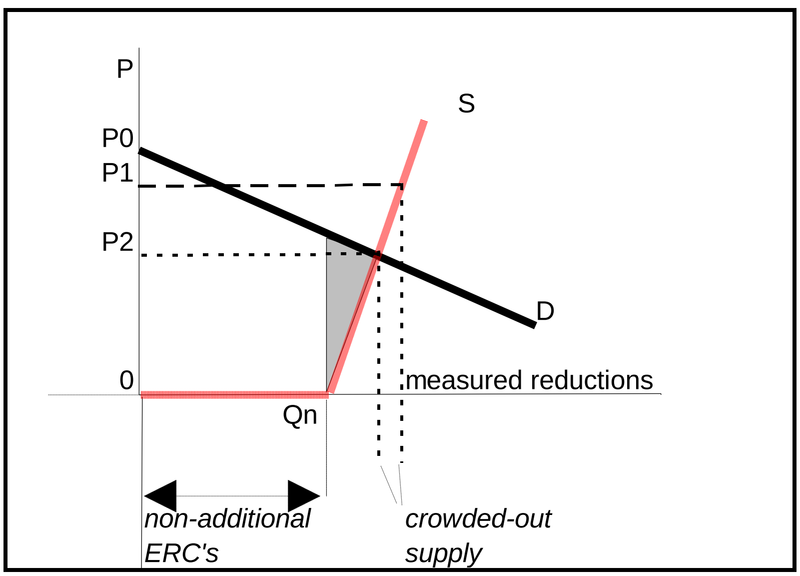

Figure 2

ERRORS IN BASELINE DETERMINATION AND THEIR COSTS

It's not possible to determine baselines with perfect precision. In the important case of determining additionality, there are two types of mistakes: certifying non-additional projects (a type II error), and denying certification to genuinely additional projects (a type I error). Each kind of mistake carries a cost. The former increases world GHG emissions; the latter denies funding to a worthwhile project and increases world abatement expenditures.

Any system for certifying ERs is bound to make one or both of these types of errors. There's likely to be a tradeoff between the errors. For instance, it's been suggested that ER eligibility could be limited to a class of projects deemed obviously additional– say, solar generation of electric power. This reduces type II errors, at the cost of massive type I errors. On the other hand, lax enforcement of eligibility requirements eliminates type I errors, but introduces serious type II errors.

Note that the errors do not, in general, cancel out. That is, we cannot apply the standard statistical argument that, if our baseline estimates are unbiased, the mean *estimated* baseline over all projects will be very close to the mean *actual* baseline. This is clear in the case of additionality determination: each type I error screens out a useful project (but does not affect net GHG emissions), and each type II error increases global GHG emissions. More generally, projects with underestimated baselines will be less financially viable and may as a result be withdrawn, or fail. Hence even if baseline estimates for project candidates are unbiased, the baselines for successful projects will tend to be biased upwards.

The magnitude of these problems depends on the relative supply of projects whose additionality is in question, and on the costs of accurately distinguishing between additional and non-additional projects.

RETROFIT/REPLACEMENT TYPE PROJECTS

Additionality determination is a particularly acute problem for energy-related projects which involve the retrofitting or replacement of inefficient apparatus. This class of projects probably includes some of the lowest-cost supply of GHG reductions. The existence of these opportunities is one of the reasons that there are perceived strong gains to ER trade between developed and developing countries. The problem is that it may be difficult to distinguish between low-cost ER projects, and "no-regret", non-additional projects. A screening system may be prone to making both type I and type II errors in this case – and the type I errors (exclusion of valid projects) may be particularly costly.

SUMMARY AND CONCLUSIONS

Strong forces will tend to favor upwardly-biased baselines, and proposal for ERs from nonadditional projects. If not screened, the result is increased world GHG emissions, reduced revenues to valid ER projects, and undermined confidence in the FCCC.

However, screening is neither costless nor perfect. Screening will result in exclusion of valid projects, as well as inclusion of invalid ones. Error rates can be reduced, but at a cost. The following sections examine the sources of inaccuracy and uncertainty in baselines and additionality, and discuss the applicability of actual and theoretical methods for estimating baselines.

3. ADDITIONALITY ISSUES

THE KEY ADDITIONALITY QUESTION: WHEN WOULD TECHNOLOGIES SHIFT?

Consider a JI project which substitutes a new, highly efficient gas boiler for a inefficient old coal boiler, and takes as its baseline the continued operation of the coal boiler for 20 years. Baseline determination has two elements:

1) What would be the emissions of the coal boiler, if it continued in operation?

This is largely, though not entirely, a monitoring or engineering issue. Engineering methods can be used either to directly measure emissions, or to relate emissions to more easily measurable proxies such as fuel consumption. This paper will largely sidestep this measurement issue. It can be solved in principle with equipment and statistics – though some systems may simply be too expensive to monitor with acceptable accuracy. Large measurement errors would result in very uncertain measurements of offsets from projects offering small emissions reductions.

2) Would the coal boiler in fact have continued in operation – and how long? Twenty years? Five years? Three months?

I will argue that this is the more important and difficult question. The short history of AIJ/JI projects provides examples of projects which appear, in retrospect, not to be additional. For instance, in Pyrzyce, Poland, a bilateral AIJ investment sponsored the replacement of 68 coal-fired boilers with a central geothermal heating plant. (Nordic Council of Ministers 1996) The baseline assumed the indefinite continued use of the individual coal-fired boilers. A review team subsequently found that the local authorities had two backup plans in the event that AIJ funding did not materialize: installation of a central, modern coal-fired plant, or installation of the geothermal plant with a lag of a couple of years.

Many types of proposed or actual JI/ER projects share the same characteristic: they sponsor a discrete technology switch which arguably might have occurred spontaneously, in the near to intermediate future. This is particularly true for retrofit/replacement type projects, but also applies to technology choices for new facilities. Some examples are as follows:

Project type	*Factors affecting spontaneous adoption of new technology*
Fuel-switching projects, especially away from coal	value of fuel savings and air pollution reductions; maintenance cost of old plant
New generator choices: low or high efficiency?	valuation of fuel savings and air pollution reduction; maintenance of fuel subsidies or price controls on electricity
Demand side management: installation of energy-saving equipment	valuation of energy savings
Install coal processing and washing improvements	price differential for processed coal
Methane capture from landfills for electric generation	Standards for landfill construction: *Landfills with minimal standards*: installation of methane capture has costs greater than benefits; therefore project is additional and abates methane. *Landfill with high standards*: much infrastructure needed for power generation is already in place, power generated from methane more than defrays investment costs, therefore project is <u>not</u> additional
Adopt reduced impact vs. standard logging techniques	Do loggers save money or satisfy regulatory requirements with low impact techniques? How strictly will the government enforce logging regulations?

TECHNOLOGY SHIFTS DEPEND ON HARD-TO-OBSERVE PARAMETERS

Reviewing this list, additionality questions arise when technology adoption decisions depend on parameters which are *hard to observe, subject to misrepresentation, subject to strategic manipulation, and subject to change.* Crucial parameters of this type are:

- cost of capital/risk premia: in many economies in transition, capital costs are very high and hard to gauge precisely. The problem is compounded for risky investments. This may be the single most important parameter affecting additionality in energy-related projects. These projects almost always involve an up-front investment which yields a stream of benefits in terms of fuel cost savings. For additionality, what matters is whether the return from this investment is sufficient to induce self- or external financing of the project.

- environmental charges and enforcement levels for SO2, NOx, and particulate emissions: Efficiency-enhancing, carbon-saving energy investments generally reduce standard air pollutants such as particulates and SO2, which impose considerable local health and economic costs. These implicit costs vary from place to place. In many cases there are explicit regulations or charges associated with air pollution; the official level is observable, but the effective enforcement levels vary considerably from place to place and over time (see Wang and Wheeler 1996 for a discussion of geographic variation in effective pollution levies in China). It is these enforcement levels which will determine the degree to which environmental benefits are weighed in the host's decision about technology changes.

- maintenance and downtime costs: where the reference project is continued use of an old facility – for instance, a half-century-old boiler – maintenance costs will increase over time, at a rate that is hard to predict.

- transactions costs – Even in well-functioning developed economies, highly profitable opportunities for energy savings are overlooked. The failure to take advantage of these opportunities is often attributed to ill-defined (but possibly real) transactions costs, though it is clear that poor incentives also play a role.

- energy prices – These of course are currently observable, but their unpredictable future changes will be a major determinant of incentives to switch technologies.

- opportunity costs of land. The profitability of switching to a carbon-friendly form of land use (such as plantations or agroforestry) depends on the profits from alternative uses, such as pasture. Average returns to these activities may be observable, but there may be a good deal of variation based on land quality, distance to market, and management skills.

- commodity and timber prices – again, these are subject to unpredictable future change and will have a large bearing on technology switches. For instance, as cattle prices decline, some areas of pasture will be spontaneously abandoned to secondary regrowth of forest.

- public policies affecting energy prices: trade, fiscal, and regulatory policies affect fuel prices and electricity tariffs and thereby influence technology choice. It is difficult to predict whether or when these policies will be modified or changed.

- public policies affecting agricultural and forest products prices – import restrictions, credit subsidies, and price supports affect decisions on the conversion of forests to agriculture.

- enforcement levels for forestry and land use regulations – many countries have laws prohibiting unauthorized forest clearance and placing strong restrictions on forest exploitation. As in the case of pollution regulations, these laws are often imperfectly enforced – but they are enforced to some degree. Over the decades-long horizon of a potential forest protection project, what assumptions should be made about enforcement levels?

  In order to verify additionality and construct a reference scenario for a JI project, we need to be able to impute or predict these difficult-to-observe parameters[2]. We run into three problems:

  1. *Actors have an incentive to hide or misrepresent these parameters.* It is to the advantage of project sponsors to be able to claim that their capital costs, risks, training and setup costs are high. These then constitute barriers to the adoption of the JI/ER project.

  2. *The parameters are subject to change over time.* This is a problem for retrofit/replacement type projects, where the baseline scenario involves continued operation of an old facility. While it may not be advantageous to shut down that facility today, it may become advantageous in the future, depending on how prices, capital costs, and pollution regulations evolve. In many cases the direction of change may be somewhat predictable. For instance, in economies in transition, there may be a trend towards lower capital costs, higher pollution charges, higher fuel charges, and higher electricity tariffs, all of which would promote switching from an old, inefficient energy technology to a new, more efficient one.

  3. *The parameters may be distorted.* This is a problem of equity across countries, and possibly of moral hazard. It occurs when countries or other actors increase (or threaten to increase) emissions, or maintain undesirable policies, that establish artificially high baselines. Possibilities include:

  - energy subsidies: subsidized fuel prices encourages the adoption or retention of low-efficiency generators, boilers, and energy distribution systems. Subsidized electricity artificially encourages low-efficiency and low-productivity end-uses.

  - deforestation subsidies: including subsidized agricultural credit and technology.

  - neglecting conservation: By restricting the size of its national system of protected areas, a country increases the area of forest available for agricultural conversion or timber exploitation and therefore available for generating ER's via protection.

---

[2] See section 4 for a discussion of how this approach relates to the idea of "barriers".

- retarding development of market infrastructure: offsets markets may retard the development of mechanisms to finance energy conservation measures, such as energy service companies, regulatory reform, and financial sector innovations.

- reducing efforts to enforce pollution and forestry laws: lax enforcement establishes a baseline of inefficient energy use and rapid deforestation against which offsets can be claimed.

It's worth stressing that this is a genuine dilemma which does not necessarily ascribe base motives to would-be host countries; the term 'moral hazard' doesn't do justice to the problem. Many countries, for instance, have strict forestry laws on the books. Enforcing them creates conflicts with powerful vested interests and often gains little public support. Officials who believe that maintenance of forest cover is socially beneficial now find that, by *not* enforcing the laws they can gain enough resources to maintain the forests while keeping both the public and the vested interests happy. On the other hand, this kind of strategic behavior may be seen as inequitable by other countries that have taken greater steps to enforce similar laws and therefore find themselves unable to claim offsets.

SOME EXAMPLES

Hidden parameters: Pyrzyce district heating

Nordic Council of Ministers (1996) presents a detailed financial and economic analysis of the Pyrzyce coal-to-geothermal project mentioned above. The project involves an investment of $15.31 million over two years. After this start-up period, it delivers estimated annual savings of $890,000 in fuel costs, $130,000 in maintenance costs, $1.97 million in value of SO_2 and NO_x reductions, and 68,618 tons of CO_2 reductions. Could this project be undertaken on a commercial basis? Drawing on the data presented in Nordic Council of Ministers (1996), the table below recalculates the net present value of costs (in millions of dollars) under different assumptions about two key parameters: the opportunity cost of capital, and the value of SO_2 and NO_x reductions (expressed as a multiple of the original assumed values). Positive numbers mean that costs exceed benefits, and suggest that under these conditions the project is truly additional. Negative

		Discount rate		
		0.05	*0.15*	*0.25*
Pollutant cost factor	*0*	$2.83	$7.51	$8.53
	0.5	($8.51)	$2.18	$5.38
	1	($19.84)	($3.15)	$2.23
	2	($42.51)	($13.82)	($4.07)

numbers, shaded and in brackets, mean that benefits exceed costs, suggesting that the project is not additional.

According to this table, if this project were located in an area which placed no value on SO_2 and NO_x reductions (cost factor=0), it would not be undertaken; even with a very low discount rate, the fuel and maintenance savings are not sufficient to compensate for the investment costs. On the other hand, a similar project located in an area with very high sensitivity to air pollution (cost factor=2) would be undertaken even under very high discount rates. At the actual pollution sensitivity (cost factor=1), additionality is very sensitive to the cost of capital; below about 19.5% the project is worthwhile, above that threshold it is not.

In most economies in transition, these two parameters – capital costs and effective pollution charges – can be estimated, but only with some uncertainty. For instance, there may be official pollution charges, but enforcement may vary systematically between regions and between types of plants. This means that outside observers might find it difficult to determine whether a Pryzyce-type investment is additional in a particular setting – say one in which the investor applied for JI finance, claiming a capital cost of 17% and a pollution cost factor of 0.8. It is for this reason that simple rules of thumb (e.g.: geothermal is always additional) are likely to be unreliable.

This analysis is consistent with the finding noted earlier that the town planned to abandon the old heating system even in the absence of AIJ funding.

Uncertain parameters: reforestation of pastures in Costa Rica

Faris *et al.* (1997) analyze the financial and economic returns of a reforestation JI project in Costa Rica. (This is a stylized version of a current project). The project involves converting pasture to timber plantations. In addition to the sale of carbon offsets, project owners benefit from a small harvest of wood at year 12 after plantation establishment, and a large one at year 20. The assumed baseline is indefinite maintenance of pasture, with a revenue of $20/year.

The additionality or baseline question in this case is whether the investors would have found it profitable to invest in the timber plantations in the absence of carbon offset sales. We reanalyze Faris *et al.*'s spreadsheet to examine the sensitivity of additionality to assumptions about the opportunity cost of capital, and to variations in wood revenues 20 years hence. Future wood revenues are subject not only to price risk, but also to risks of damage or expropriation.

The net present value of costs/hectare are shown below as a function of capital cost, and of year 20 revenue expressed as a multiple of the original assumption. Again negative costs (shaded) indicate a profitable project. On this analysis, risk appears to be an important determinant of additionality. Assuming that the wood revenue factor has an expected value of 1, the project is fairly attractive at 6%. At an 8% discount rate, the project begins to look unattractive to a risk-averse investor. As the discount rate rises above 10%, the project is unattractive to a risk-neutral investor, suggesting that it is additional.

	NPV of COSTS under different scenarios
	discount rate or cost of capital

		0.06	0.08	0.10	0.12	0.14
Wood revenue factor (year 20)	0.25	496.73	618.67	688.66	724.97	739.49
	0.50	26.81	301.32	472.79	577.11	637.52
	0.75	(443.10)	(16.03)	256.92	429.24	535.56
	1.00	(913.01)	(333.38)	41.04	281.38	433.60
	1.25	(1,382.93)	(650.74)	(174.83)	133.52	331.64
	1.50	(1,852.84)	(968.09)	(390.70)	(14.35)	229.68

Distorted parameters: policies and emissions in Indian coal plants

Khanna (1997) models emissions and electricity production at 63 coal-based power plants in India, accounting for 86 percent of coal-based generating capacity in 1990-91. About half these plants are operated at energy efficiencies of less than 25%, against design efficiencies of 32%. Several policy-related factors keep efficiencies low and CO2 emission high. Most plants use low-quality coal, because imports of washed coal are prohibited. Subsidies for electricity production, and fixed electricity tariffs, keep low-efficiency plants in operation.

Khanna shows that removal of trade restrictions, subsidies, and price caps results in an annual welfare gain of $600 million, a reduction in government subsidies of $3 billion, and a reduction in CO2 emissions of 11.7 million tons.

If JI projects were proposed in this sector, what would be the appropriate baseline? It depends whether policies are taken to be mutable. One could argue that it is politically impossible, in the short to medium run, to remove the large subsidies. In this case, the current situation would be accepted as the baseline, and ER sales could be used to finance CO2 reductions. Alternatively, one could argue that the country has the power unilaterally to reduce emissions, and should not be rewarded for maintaining socially inefficient policies. In that case, the baseline would be drawn 11.7 million tons under current emissions. It's worth noting, though, that in this case the volume of ER sales could not possibly finance the generation subsidies, so the ER market does not provide a perverse incentive (to the government) to maintain them.

SUMMARY AND CONCLUSION

The hard part of baseline setting is determining whether, or when, the project sponsor would have spontaneously switched from a high carbon to a low carbon activity. The switching decision depends on hard-to-observe parameters. To construct a baseline, we can either impute those parameters and predict the sponsor's behavior, or look for a control group which represents the baseline conditions. The next section examines both approaches.

4. METHODS FOR DETERMINING ADDITIONALITY AND BASELINES

INTRODUCTION

This section describes four approaches to baseline determination, focusing particularly on determining whether or not the proposed low-GHG project would have been spontaneously adopted:

1. *Direct questioning:* Ask the project participant what would have been done absent the project.

2. *Control group methods:* Observe the behavior of a comparison group not offered opportunities to sell ER's.

3. *Behavioral/financial models:* Build a model that predicts how a facility or sector would respond to the incentives posed by an ER project. A powerful methodology for building such a model is to assume that the actors maximize profits, subject to some constraints: "should have" as a means of forecasting "would have".

4. *Sectoral or regional cap:* This redefines a JI project as the establishment of an allowance-like system, where unused allowances are equivalent to offsets.

Of these, the control group and behavioral/financial models are likely to be the most useful.

DIRECT QUESTIONING

The most straightforward approach to baseline determination is to ask the project sponsor what would be done in the absence of the project. To a skeptic, this approach will seem hopelessly naive and open to manipulation, but is widely used. Interestingly, it is the main way that US utilities estimate 'free ridership' when assessing the impact of demand-side management programs. (See section 7).

DSM programs offer incentives to households or firms that install energy saving measures. Utilities are rewarded for energy savings after correcting for 'free riders' - those who would have installed the measures anyway. To determine the free ridership rate, evaluators often use survey instruments such as that shown in Box 1. Potential drawbacks of this approach include respondents' inability to deal with hypothetical questions (Ozog and Waldman, 1992), their incentive to answer strategically, and their reluctance to admit to 'free riding'. Remarkably, a significant proportion of the respondents acknowledge that they would have adopted the measures without any incentives.

More sophisticated surveys aimed at large commercial customers use a wider range of questions to establish decision processes and rules. These are similar to the behavioral models discussed later in this section. For instance, Goldberg and Scheuerman (1997) describe evaluation of a program in which commercial customers were provided with incentives to adopt high-efficiency lighting. An evaluation survey found that lighting technology decisions were often determined by formal corporate policies established at distant headquarters, rather than by local incentives; on this basis, free ridership was estimated at 49%.

These methods are now being applied to DSM projects in developing countries. It is worthwhile to monitor these efforts and assess the potential applicability to JI/ER projects that similarly consist of 'bundles' of household or firm-level interventions.

The following questions appear in a survey of residential participants in a utility-sponsored energy audit program. These two questions are repeated for each of five types of installation which might have been recommended to the customer in the course of the audit.

Q. **39. On a scale of 1-10 with 1 being "definitely would not have installed" and 10 being "definitely would have installed", how likely would you have been to install this measure on your own if it had not been recommended to you through the Audit**?

1	2	3	4	5	6	7	8	9	10
Definitely would not have (*skip next question)*									Definitely would have

98 DON'T KNOW... *skip next question*

Q. **40. When would you have installed this measure?**

1. Within one month of scheduling audit
2. Within six months of scheduling audit
3. Within one year of scheduling audit
4. Over one year from time of scheduling audit
8. DON'T KNOW

Source: Hagler-Bailly

Box 1

CONTROL GROUPS

A valid control group is the gold standard of baseline determination. For instance, if high-efficiency light bulbs were subsidized through a JI/ER project in one city, but not in an otherwise completely comparable control city, monitoring the latter would provide baseline information about the spontaneous rate of adoption of the bulbs in the absence of incentives.

Valid control groups are also the holy grail of baseline determination, because they are difficult to find. This section begins by reviewing the pitfalls associated with a control group approach, then discusses some potential solutions, and concludes with a discussion of the applicability of the control group approach to baselining for various ER project types.

Why valid control groups are hard to find

Ideally, we would like to compare the behavior of a 'treatment' group or individual offered the opportunity to participate in a class of ER projects, with a control group not eligible to participate. After controlling for compositional differences between the groups, and differences between groups in exposure to exogenous factors (such as weather), the control group provides a baseline against which emissions reductions can be reckoned.

Two practical problems stand in the way of this straightforward approach:

Idiosyncrasy: Valid statistical comparisons require a decent sample size to detect modest changes in emissions; much will depend, of course, on the degree of noise and confounding variation in the data. In many cases, the project facility may be unusual or idiosyncratic, and it may be difficult to find a large enough or similar enough control group to permit these comparisons. This will particularly be the case for fuel-switching projects involving large industrial or municipal facilities.

It will also, in general, be the case for evaluating projects affecting national electrical generating capacity. This is because any such project has *sector-wide* impacts. This point is made by Swaminthan and Fankhauser (1997), who illustrate how the alternative to installing a small renewable-energy plant is to accelerate the phase-in of large conventional plants. The unit of analysis for project-vs.-control comparisons is therefore not the plant or utility, but the entire national generating sector, so there can be no domestic comparison group.

Selection effects: The immediate problem for ER applications is that all potential ER suppliers within a participating country are potentially eligible to participate in the offsets markets. Would-be project sponsors will systematically recruit potential offset suppliers with the interest and capability of cost-effectively reducing emissions. As more and more suppliers are recruited into projects, remaining nonparticipants constitute an increasingly less appropriate control group, since they are likely to be systematically different from their counterparts who were recruited into a project. In the limit, there may be no nonparticipants left at all.

Before and after comparisons

The simplest possible control group approach is a before-and-after comparison: emissions of the entire eligible population (of facilities, forest plots, etc.) are compared before and after the advent of the ER project. Given a long enough time series, and sufficient variation, the before-and-after comparison can be adjusted for possibly confounding exogenous factors such as weather.

There are several drawbacks associated with before-and-after comparisons:

- *Moral hazard* is a danger. For instance, if pre-program deforestation rates serve as a baseline, there is a danger of inducing higher deforestation in any area which might later seek to produce offsets through deforestation prevention. A standard corrective in this situation is to use older data – say, before 1995 – to establish the baseline, but this raises the second problem: failure to control for contemporary trends.
- *Changing incentives or conditions* make historical data a poor guide to the current baseline. For instance, in the economies in transition, rapid changes in prices, management structure, regulations make historical comparisons of limited use. Similarly, past deforestation rates may be of little use if subsidies for forest conversion or prices of agricultural products have changed substantially, and may change in the future.
- *Recently-introduced technologies* may still be in the process of natural diffusion; typically adoption rates follow a logistic curve in the period after an innovation is introduced. A simple before-and-after comparison would overstate the baseline if it failed to account for natural diffusion rates. If it were possible carefully to model the diffusion process, however, before-and-after comparisons could in principle detect 'spillover' effects of ER projects on inducing technology adoption by nonparticipants.

Comparing project participants and nonparticipants over time: self-selection and other problems

The alternative to before-and-after comparisons is concurrent, post-project comparisons of 'treatment' (project) and control groups. Concurrent control groups account for ongoing changes in the economic environment, but present their own problems. First, the use of concurrent control groups implies the use of dynamic baselines – baselines that are not prespecified, but 'observed' in the course of project execution. While often regarded as infeasible, this kind of dynamic baseline has been used routinely for the determination of DSM incentives for net energy savings (see section 7).

Concurrent treatment-vs.-control comparisons are straightforward when the control group is ineligible for, and likely to be unaffected by, the project. In many, perhaps most, cases these conditions will not be satisfied. The 'control' group will consist of units (households, firms, etc.) which could have participated in the project, but chose not to, or were purposefully not offered the opportunity. If project participants constitute a significant fraction of the universe of possible participants, then they will tend to differ systematically from nonparticipants. It's generally reasonable to suppose that participants may have been more willing to reduce emissions even in the absence of a project, meaning that the baseline is overstated.

For instance, consider a program which offers incentives for industrial firms to adopt high-efficiency lighting. Acceptors of the incentives will tend to have lower discount rates, greater internal incentives for cost-minimization, and higher maintenance costs than nonacceptors. These characteristics would be conducive to energy conservation and emissions reductions even in the absence of incentives. But, as emphasized in section 0, these characteristics are difficult to observe. Therefore non-acceptors are a biased control group – they don't really represent how the *acceptors* would have behaved, absent the incentive.

This problem has long been recognized in DSM applications (EPRI, 1991) and indeed is the canonical problem of the program evaluation literature: comparisons between control and program groups need to control for observed and unobserved confounding variables. Train (1994) presents a detailed description of the econometric issues and a critique of the state of practice.

A simplified version of the approach is as follows. Assume that an ER project offers units (firms or households) an opportunity to participate in a program – perhaps one which involves incentive payments. The program promote measures which reduce emissions. It is quite possible, however, for nonparticipants to adopt these measures, and this leads to a free-rider or additionality problem in determining the baseline.

Behavior of the units can be described through a system of equations:

$$P^* = X\beta + u \qquad (1)$$

$$P = 1 \text{ if } P^*>0,\ P=0 \text{ if } P^*<0$$

$$\Delta E = Z\gamma + \delta P + e \qquad (2)$$

where P^* is the unobserved propensity to join the program

P is a dummy variable indicating participation in the program

X and Z are vectors of variables affecting participation and emissions reduction

ΔE is the observed change in emissions after the project was initiated.

Program participation is based in part on the unit's predisposition to adopt the measure and save energy. Therefore the observed determinants of emissions reduction, Z, overlap substantially with the determinants of participation, X; and likewise for the effects of the unobserved determinants u and e. This means that the observed participation indicator P is highly correlated with the unobserved propensity to reduce emissions e; in the limit, if everyone who was going to reduce emissions anyway volunteers to participate in the program, then the true value of δ is 0, but an OLS estimate of equation (2) yields a spurious positive value.

The well-known econometric solutions to this problem are:

a) to estimate the pair of equations jointly via maximum likelihood methods, allowing for the correlation of e and u; or

b) to estimate the participation equation first, use it to derive E(u|X) (the expected unobserved propensity to participate, given observed variables X), and then to plug this value into equation (2) as a control for the self-selection bias.

This cookbook procedure faces some difficulties in implementation. Train (1994) notes that this procedure is often erroneously applied to the determination of the *level* of energy consumption (or emissions) E, rather than the *change* ΔE. In addition, the econometric identification of equation (2) is tenuous. Intuitively, to be sure that we are capturing the exogenous effect of project participation on emissions reductions, we need to find a variable which affects participation, but which does *not* affect emissions reductions. Again, intuitively, just about the only conceivable candidates for this role are variables which describe bureaucratically-determined eligibility requirements or recruitment/advertising efforts which vary between units.

International comparison groups

When should international control groups be used? In principle, the use of international comparisons provides a means of providing true, uncontaminated control groups not subject to self-selection bias. However, it may often be the case that international comparison groups differ too much in composition, behavior, or ambient price levels to make satisfactory controls. This will be particularly true for situation involving retrofits of old energy facilities, or for forestry applications. International comparisons may be more reasonable with respect to the adoption of new technologies -- such as electric generation equipment – where the menu of possibilities is indeed fairly standard throughout the world. Even here, though, differences in capital costs may complicate the comparisons.

There are two situations in which the use of international comparison groups is advantageous, even if the result is to demonstrate the lack of opportunity to generate ERs. The first is where domestic prices are distorted. Here international comparison is an important means of determining the baseline in an undistorted environment. The second is in the case of footloose industries selling to a world market. Some fear that firms in these industries could claim offsets without actually reducing emissions, merely by relocating their facilities from the developed to the developing world and then using local firms to define the baseline. Use of international comparators for this class of projects would reduce this risk.

Summary: applicability of control group approaches

In sum, control group approaches are most useful when:

- the number of project and nonproject observation units (e.g., firms, households) is large
- the units are reasonably homogenous, or their emissions behavior is easily related to well-observed characteristics

- the project is of limited geographical scope or otherwise does not recruit most of the pool of potential participants (this criterion can at best be satisfied only temporarily, since there is no obvious restriction on participation in ER markets)
- there are no large domestic policy distortions such as subsidized energy prices
- spillover effects are not large, or can be separately modeled.

Preproject data can serve as a useful comparison basis when factors affecting emissions don't change much over time. However, the routine use of pre-project data as a baseline could lead to moral hazard for subsequent projects, as would-be participants seek to establish a higher baseline.

Concurrent control vs. project comparisons are potentially useful in controlling for unpredictable factors such as weather, capacity utilization due to business vagaries, and prices. However, use of concurrent data implies a dynamic baseline.

These characteristics make the control group approach to baselining applicable to many DSM projects. As section 7 discusses at length, control group techniques of varying rigor have been extensively used in the evaluation of net energy savings from these projects in the US. Similarly, control groups may be an effective way of establishing baselines for fuel-switching projects involving large collections of small or medium-sized facilities. They may be effective also for projects involving the adoption of reduced-impact logging or other agricultural/silvicultural technologies among small and medium operators.

In general, control group techniques will be less satisfactory for certain important classes of projects:

- *large retrofit or replacement projects in transition economies*. Here the baseline question is when the original equipment would have been replaced, absent the project. Historical data will provide no guide, and the universe of contemporaneous control facilities may be too small.
- *large electrical generation projects*. As noted above, the need to look at sector-wide impacts makes these projects awkward for control group analysis.
- *deforestation prevention or reforestation projects*. The baseline, without-project rates of deforestation or regrowth depend strongly on prevailing prices for wood and agricultural goods and on the state of enforcement of forestry laws. These parameter may well vary over the 20 or 30 year period that might be typical for such projects. It may also be impractical to set up a control area which is both excluded from participation in the project and insulated from the project's indirect effects (such as a geographical displacement of the demand for land conversion).
- *projects in host countries with energy subsidies or other policy distortions.* If it is determined that baselines should be based on economic prices rather than prevailing, subsidized prices, local control groups are not helpful.

These shortcomings motivate the next approach to baseline determination.

BEHAVIORAL MODELS BASED ON FINANCIAL/ENGINEERING ANALYSES

In the absence of a real control group, it is necessary to create a virtual control group. This means constructing a model describing how the unit in question would behave over time in the absence of offset sales. The focus would be on predicting whether the unit would adopt the project or the reference (baseline) technology, in the absence of a market for GHG offsets.

Such a model would have both normative and positive elements. It would be normative to the extent that it corrects for policy distortions – for instance, by shadow pricing energy at world prices. It would be positive to the extent that it attempts to predict how the unit would actually behave, contingent on those corrected market signals. Such a model would not, for instance, assume that firms face no transactions costs, no risks, and can borrow at the social discount rate. It would recognize that there is extensive evidence that firms do not invest in *apparently* high return energy measures (see DeCanio and Watkins 1998 for an empirical study and citations to the literature) – but explicitly model this as a consequence of high transactions costs and high capital costs.

The approach

One approach would employ financial (cost/benefit) analysis of engineering (or agronomic) models. This approach would use the same model, and level of sophistication, as would be employed to make an investment decision in the project. It would in fact simulate the investment decision. The proposed procedure is as follows:

1) Construct a cash flow model which predicts project costs and benefits over time, as a function of output prices, input prices, and important contingencies.

2) Use the model to evaluate potential possible projects, including the JI/ ER project (evaluated without possibility of ER sales), and one or more reference projects.

3) On the basis of (2), predict which project would be chosen in the absence of ER sales opportunities, using a normative investment decision rule. (A positive rule might be better but would be harder to derive.) For instance, in a retrofit/replacement project, the rule might be: invest in the ER if it offers a greater net present value than continued operation (with optimal maintenance) of the current equipment. More complex, heuristic models of the investment process are possible.

The emphasis on financial analysis subsumes and makes more rigorous an alternative approach, which is to identify "barriers" to adoption of the project technology. (IEA 1997; Carter 1997) There are many plausible barriers, including poorly functioning financial markets, risks associated with installing and operating locally unknown technology, and internal organizational structures that discourage investments in energy efficiency. (See for instance Golove and Eto, 1996). However, the critical assumption behind JI or ER projects is that these barriers can be overcome, given enough money. Money overcomes barriers either by covering unusually high transactions or set-up costs, or by boosting returns high enough to compensate investors or lenders for the project's risk. A financial or behavioral analysis provides a framework for quantifying the effect of these barriers on costs, risks, and returns. It therefore provides a systematic framework for assessing additionality claims.

This approach is similar to the incremental cost analysis required by the GEF. The GEF finances incremental costs of a project relative to a baseline scenario. However, the GEF incremental cost analysis is made awkward by the explicit exclusion of incremental benefits. Analysts are urged to find some way of describing incremental benefits as avoided incremental costs (GEF 1996, para 25). For the purposes of analyzing JI/ER projects, there need be no embarrassment at explicitly factoring incremental benefits into the investment decision framework.

The need for default parameters

Section 3 argued that behavior depends on some hard-to-observe parameters – in particular, the firm's capital cost or target rate of return, and the penalties attached to air pollution. To calibrate any behavioral model, we have to specify those parameters.

To facilitate baseline determination, baseline certifiers could agree on standard, country-specific default values for these crucial but unobservable parameters, including cost of capital (or target rate of return), and effective pollution charges. Standard values could also be used for current and anticipated energy prices. For instance, analyses of fuel-switching projects would be required to use common values for energy prices, actual or shadow prices for pollution, and target rate of return. These values would necessarily be country-specific. Different rates of return would likely be set for new versus established technologies.

The use of standard values for these parameters has two advantages. It simplifies project preparation by obviating the need for researching and justifying these contentious values. It also removes one of the chief levers for 'gaming' the system.

The use of standard values would, of course, introduce type I and II errors, as described in section 2. For instance, if the risk-adjusted target rate of return is set as 22%, firms with lower capital costs will tend to have nonadditional projects approved; firms with higher capital costs will tend to have valid projects rejected. Error rates could possibly be reduced by tying the default values to firm size, type, or location, but this makes the process more complex.

A major policy decision is whether to use these default parameters to correct for policy distortions. Should, for instance, the calculations be done at world energy prices or at prevailing local prices? Where there are no effective pollution charges, should a shadow price of pollution be imposed? Should official pollution charges be used if they are not actively enforced? These questions boil down to the decision: should a country's policies be taken as mutable or given? (See section 3). Ultimately there may be a clarification of the Kyoto Protocol which decides this issue. Until then, practitioners will have to decide on their own. It would however be possible for an ER project to calculate offsets against two baselines (corrected and uncorrected for policy distortions), pending a formal decision under the UNFCCC. It is worth noting that the GEF methodology for computing incremental costs requires that the calculations be done at world prices if there are local distortions. (GEF 1996).

New technologies, risk, and diffusion

Technologies new to a country would be expected to have higher start-up costs and higher risks of failure. In the evaluation procedure, this might be reflected as a higher target rate of return. It will be hard to set this hurdle with precision. Setting the hurdle very high (i.e., favoring a presumption of additionality) may be relatively harmless in this case because of the high likelihood of positive spillovers. As the technology diffuses, relative risks and start-up costs will decline and it will be possible to model the adoption decision with greater accuracy.

Incentives for accurate reporting

Couldn't a project sponsor manipulate financial or engineering records to support an additionality claim? Of course. The advantage of the proposed financial analysis, as with any auditing system, is that it makes manipulation more difficult.

It's possible to structure incentives to promote accurate reporting. One way to do so is to employ a partial-crediting system (see section 0) for ERs as a Bayesian crediting mechanism: we credit the *expected* number of genuine ER's. The rationale here is that there is genuine uncertainty, for the host as well as the certifier, about the costs and benefits of both the project and reference scenarios. Again consider a typical retrofit/replacement project. A Bayesian approach to additionality might compute the internal rate of return of adopting the ER project, relative to maintaining the status quo. If the IRR falls below some minimum threshold, the project is presumed to be additional with 100% probability. If the IRR falls above some maximum threshold, it is presumed nonadditional with 100% probability. Between the thresholds, the presumed probability of the project's non-adoption is interpolated. The gross amount of reductions produced by the project is scaled by this probability. This sliding-scale approach reduces the strong incentive to manipulate data that would result from an all-or-nothing determination of additionality.

Proper incentives could be further reinforced in this case if there were a mechanism for funding high-return energy-efficiency projects. Suppose, for instance, that a funding mechanism, or energy service companies, were prepared to invest in energy conservation measures with payback periods of two years or less. We might then credit ER projects on a sliding scale, with full crediting for projects with estimated paybacks of four years or more, and no crediting for those with projected paybacks of two years or less. Companies with profitable energy conservation opportunities might find it more expedient directly to finance conservation measures than to go through an ER certification process which might yield relatively few credits.

Example: Poland Coal-to-gas AIJ evaluation

Incremental cost analysis was used in the appraisal of the Poland Coal-to-Gas AIJ project (GEF 1994), which sponsored the conversion of numerous small coal boilers to gas. The approach serves as a model for the investment-decision approach proposed here. (The procedure was embodied in a "user-friendly, menu-driven spreadsheet model"). In appraisal, it was assumed that the target rate of return (cost of capital) is 25%, and used official pollution charges of $73/ton for SO2 and NOx and $36/ton for particulates. A simple spreadsheet-type cost-benefit analysis of the profitability of gas conversion for a pilot facility showed that it was more profitable for the facility owners to retain the existing coal boilers than to replace them with gas. The analysis recognized that this decision might be sensitive to future price changes. Sensitivity analysis showed that if the pollution charges were increased severalfold to reflect actual damages, if real labor costs were assumed to increase at 5% annually, and if energy prices were set at world level, the project would have an IRR of 22%, still below the threshold. This then justified the assumption of additionality.

Since I am proposing that this modeling approach is a crude but serviceable description of actual behavior, it is of interest to know how well it predicts fuel-switching by district heating plants. In fact, since the project was appraised in 1993, a spurt of similar conversions has spontaneously been undertaken in Poland[3]. Many of these conversions are funded by grants or concessional loans through the National Environmental Fund, the Ecofund, and the Bank for Environmental Protection. In addition, some district heating companies are undertaking conversions with self-financing or conversion loans. The Krakow district heating company, for instance, is converting or eliminating about 80-100 coal boilers per year[4].

Does this trend invalidate the use of the simple investment model? Probably not. It is likely that these conversions are explicable by an unexpectedly rapid decline in the risk-adjusted cost of capital, and by an increase in the availability of concessional funds from national sources[5]. It may however indicate that the model is too simple and requires the refinement.

---

[3] Eric Martinot, personal communication, Oct. 27, 1997.

[4] *Ibid.*

[5] The latter raises very profound baselining questions. If Poland was willing to subsidize such projects on the basis of their local environmental benefits, that might well be taken to show that such a project could not possibly be addtional, and qualify as a source of emissions reductions. But if we agree to assess projects at world, rather than local, fuel prices, should we do the same thing for pollution charges?

Since district heating projects have been, and are likely to continue to be, a prominent class of JI/ER projects, it would be worthwhile to test the accuracy of alternative financial/behavioral models of fuel-switching. This can be done by analyzing the actual conversion experience of district heating plants in Poland and other countries over the past five years. A particularly strong hypothesis would be that concessional financing was directed towards facilities with strong local environmental impacts.

Dynamic or static baselines?

The financial/behavioral approach to modeling makes it possible to construct dynamic baselines. Consider, for instance, an old district heating plant. Based on today's fuel prices, labor costs, and capital costs, it may not be profitable for the plant to switch fuels. This would justify the creation of ER's by switching to a more efficient fuel. But how long should we imagine that the old plant continues, in the reference (baseline) scenario? It depends on how those prices are anticipated to change.

A conservative approach would presume that all those factors are systematically changing so as to favor fuel-switching. Static baseline determination would therefore:

1. predict prices, interest rates, and pollution charges over the project lifetime
2. apply the financial model to determine at what date the plant would switch fuels, in the absence of ER revenues.

The result might be, for instance, an *a priori* prediction that the plant would be retired after five years in the baseline case, so that ERs could only be generated during that period. Suppose, however, that there was some chance that prices might not change. In that event, the reference plant might continue in operation for many years. A static determination of a five year baseline would squelch the creation of many ER's.

To prevent this, a dynamic baseline could be used. The baseline would not be determined in advance. Instead, the behavioral/financial model would be exercised each year. If it predicted that the old technology would still be in place, ER's could be reckoned against this high baseline. If it predicted that incentives now favored a shift to a new technology, the baseline would be appropriately ratcheted down. This approach is more complicated than the static approach. Bear in mind, though, that actual emissions have to be measured and certified at a regular basis, so that the trouble of recomputing the baseline is not as great as might be thought.

Dynamic baselines have often been viewed as a needless, risk-increasing complication. But this need not be the case. Dynamic baselines could be tied to easily observable variables such as load factors, exchange rates, central bank interest rates, or fuel prices. They can actually reduce risk or increase the attractiveness of a project. For instance, imagine a heating/cooling project which reliably reduces emissions by 20% in a context where emissions depend on the weather. A static baseline, based on expected temperature would yield a volatile stream of offsets: lots or none, depending on how actual temperature compared to assumed temperature. Dynamic baselines can be viewed as an element of the methodological toolkit to be used when appropriate.

Summary: project types appropriate for investment decision analysis.

The financial/behavioral model approach is appropriate for use at three different scales:

1. At the *sectoral* scale, simulation models are available to determine investment in and dispatching of electric generation capacity. (See e.g. Swisher *et al.* 1997) These models could be used to determine baseline investment decisions in the absence of JI/ER projects. (Again see Swaminathan and Fankhauser 1997). Similarly, landscape-level land use models, integrated with agricultural/silvicultural supply and demand models, could be used to project emissions from land use change in the absence of JI/ER interventions. The cost of these models would be moderate to high, but would be reasonable in light of the volume of ER's produced by a sectoral project.

2. For *large projects*, the approach would mimic the investment decision methodology which would be used even in the absence of JI/ER opportunities. Examples include:

- fuel-switching retrofit/replacement projects
- choice of generator, or manufacturing technology, from among a set of "off-the-shelf" models, given a predetermined load or capacity
- decision on whether or not to build a privately-owned generating facility using renewable energy sources
- decisions by large logging companies on whether or not to adopt reduced-impact logging techniques

3. For *projects which induce changes among households or small firms*, and where control group methods are not possible, hybrid statistical/financial approaches would be used to predict behavior in the absence of the project. Examples include:

- adoption of longer-payback energy conservation measures by small households and firms
- pasture abandonment by small farmers

This proposed approach – financial-engineering models with prespecified parameters – is hardly free from ambiguity or opportunities for manipulation. However, there will be many circumstances for which there is no feasible alternative to this approach. Provision of guidelines on standard parameters and application of standard investment-decision methodologies, together with provisions to safeguard the integrity of the certification process, will go a long way towards maintaining its credibility.

SECTORAL CAPS AS BASELINES

A conceptually simple, politically difficult solution to the baseline problem is to establish sectoral or national caps and measure offsets against these. (see e.g. Carter 1997). This is particularly appealing when facility-level projects have significant sectoral effects. For instance, as noted in a couple of contexts in this paper, a decision to build a generating plant can affect grid-wide expansion and generation plans. Similarly, project-based efforts to protect a forest plot from subsistence-oriented conversion may merely divert the convertors to another location. For both energy and forestry projects, it is therefore desirable if not essential to compute sectoral level baselines and look at sectoral level effects.

There are two severe difficulties in pursuing this approach. The first is setting the overall cap. This could be done through the use of a complex model of the energy sector or of land use. It could be done on the basis prior emissions levels, adjusted for population or economic growth. In general, agreement on such a cap might be very difficult. For Annex I countries, though, the cap is already defined on a national basis and it might therefore be possible to define a sectoral subcap.

The second difficulty is allocating the rights to create offsets against this cap. The economist's natural tendency is to recommend the creation of a tradable domestic allowance system. Unused allowances would automatically count as offsets. Palmisano (n.d.) discusses the severe political problems involved in coming up with an acceptable means for allocation. ELI (1997a) acknowledges those problems, but suggests that solving them in the political rather than bureaucratic arena, and placing a legislative deadline for achieving an allocation, can advance the process.

A 'back-door' route to this system is to establish a sectoral cap, and then allow firms within the sector to generate offsets against their historically-established emissions level. This has been done in several US states in the form of Emission Reduction Credit (ERC) trading programs for NOx and VOC (volatile organic compounds)[6]. These programs allow sources to sell, for credit, reductions in emissions against a baseline. The sources already face individual limits on emissions. ERC's can be used as offsets to help satisfy these limits. Because the ERC programs take place in areas with regional limits on emissions, they strongly resemble Article 6 emissions reduction regimes in an Annex I country with a binding national emissions cap.

By and large, the methodology for baseline determination has not been contentious in these programs: baselines are specified to be the lower of permitted and actual emissions. (In some cases, plant shutdowns are excluded as a source of ERC's, for fear of leakage). The sources are already subject to regulation and monitoring, generating the information needed for baseline definition.

---

[6] The following discussion draws on the background paper, ELI (1997b).

However, the ERC systems have been criticized as requiring higher transactions costs than the cap-and-trade systems they so closely resemble. Most new ERC systems (discrete emissions reductions or DERs) require year-by-year crediting of achieved reductions against pre-established baselines. To ensure the integrity of the system, the states require three to five separate, public reports, including: a notice of DER generation, a notice of intent to use DERs for compliance purposes, a notice and certification of DER use, a notice of transfer (if the DER is sold) and a notice of DER certification. Some systems require precertification of the credits, which introduces high transactions costs; others place liability for credit validity on the buyer, which raises risks. One state uses private third-party certification. The need to verify physical output in order to trade ERCs contrasts sharply with the need only to verify the validity of an allowance certificate in a cap and trade system. (ELI 1997b; Dudeck 1995).

Can ERC/DERC systems evolve into cap-and-trade systems? Typically, ERC systems already have much of the necessary market infrastructure in place, including regional emissions caps and firm-level monitoring of output. Because of this, there are pressures for the US ERC systems to evolve into cap-and-trade systems similar to the existing SO2 system; the EPA is drafting a cap-and-trade program for NOx to cover 22 states and the District of Columbia. Palmisano (n.d.) is nonetheless pessimistic that ERC systems can evolve into allowance systems, though granting it as a long-term possibility.

It is possible that ERC systems may be suitable for EIT's faced with binding caps, especially where there is already some regulatory infrastructure in place. It would be essential to control against leakages into uncovered or unregulated sectors (such as small firms.) Over the medium run, ERC systems might facilitate a transition from JI to a pure emissions-allowance trading system for these countries

5. PARTIAL-CREDITING AND MENU CHOICE STRATEGIES

Uniform partial crediting of offsets

Suppose that we believe that there is a 50% error rate in the certification process, so that 50% of approved reductions are not in fact additional. Our problem is that we cannot identify which is the offending 50%. One response to upwardly-biased baselines is to grant only partial credits for reported reductions. For instance, we could offer 50% credit for each reported reduction, in the hope that, in aggregate, credited reductions will be about the same magnitude as actual reductions.

This partial crediting strategy has a drawback, however. Consider the situation shown in Figure 3, which shows aggregate demand and supply for ER's. The initial demand for ER's, D_0, is shown as highly elastic; think of it as being demand as seen by a price-taking small country. One half of the measured reductions (Qn) are not additional, and thus have a supply price of zero. A proposed solution is to impose a 50% in-kind tax on the ER's: buyers are required to retire, or donate to the common good, 50% of the measured reductions that they buy. This shifts the demand curve for pre-discount reductions down by 50% to D_1, because the buyers need to buy twice as many credits to accomplish a reduction which they could accomplish by other means at a cost of D_0. The partial-crediting strategy is successful in reducing both the number of non-additional credits, and the rents received by their producers: both fall by half. However, the strategy has the disadvantage of pricing out of the market some genuine, but higher-cost suppliers of ER's, those between Q_0 and Q_1. This illustrates the error trade-off discussed earlier: type II errors decrease, but type I errors increase. Moreover, the result is a kind of adverse selection: the proportion of realized credits which are not additional increases from half to two-thirds!

The partial crediting strategy can be improved if there is a means of discriminating among suppliers and reducing the rents received by suppliers with overstated baselines. We turn to two general techniques for doing this.

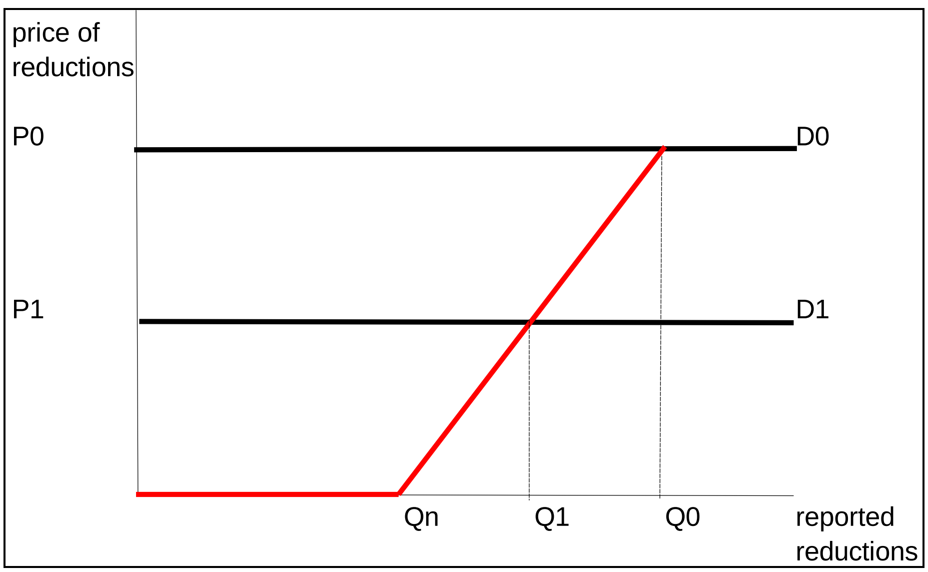

Figure 3

Relating partial crediting to rigor of baseline determination

Preparing detailed baseline analyses can be expensive, and involves some fixed costs. Small suppliers of ER's may find that the costs associated with preparing a rigorous baseline are high enough to undercut the viability of the project. But relaxation of baseline standards might invite overstatement of baselines by suppliers.

An obvious response is to offer full credit for reductions from projects presenting rigorous baseline calculations, and partial credits for reductions from projects with less-rigorous baselines meeting a minimum standard. The standards for full crediting, and the rate of crediting, will determine once again a tradeoff between type I and type II errors. Ultimately the determination has to be made partially through guesswork.

This strategy has been used in the EPA Conservation and Verification Protocols (EPA 1995; see discussions in Hagler-Bailly 1998 and Vine and Sathaye 1997). The Protocols are used for allocating SO2 allowances to utilities that encourage their customers to install conservation measures. They allow the utilities a choice among methods for establishing net energy savings. The *monitored* method requires the utilities to use comparison group methods to establish net energy savings in the first and third years after measure installation. These savings may then be applied throughout the measure's estimated lifetime, which ranges from five years for water faucet aerators to 25 years for wall insulation. An *inspection* method requires only that the utility verify that the measure remains in place. This permits 75% credit of stipulated savings rates, for 75% of the estimated lifetime. A *default* method requires no inspection and grants 50% credit for 50% of the estimated lifetime.

Revelation mechanisms[7]

Background

Regulatory economics has long faced a problem similar to baseline determination and related to the 'hidden parameters' problem of section 3. Regulators want to allow a regulated monopoly to achieve a set rate of return, but they are hampered by ignorance of the firm's technical efficiency in producing output. Under certain conditions, regulators can draw up a menu of different payment schemes corresponding to different reported efficiency levels, in such a way that the firm is induced to truthfully report its efficiency level. (Baron 1989) It is important to note that these mechanisms are not 'free'. Firms still receive information rents – payments that are larger than would be necessary in a world where their efficiency was perfectly observable.

Lewis (1997) describes how such a mechanism applied by regulators in California and other states to set payment schedules related to demand side management activities. Here the baseline was taken as given. The problem was how much the PUC should pay for net energy reductions, acting as a discriminating monopsonist on behalf of public ratepayers.

---

[7] This section is based on background material from Tracy Lewis, who proposed the menu-choice mechanism described here.

At issue was the extent and cost of potential reductions which could be achieved. Some utilities, for instance, argued that undertaking DSM programs, as mandated, would require large fixed costs and yield modest results.

The solution to the problem was the design of a menu of alternative payment options. One option offered a high fixed payment for undertaking the program, and a low marginal payment per unit of energy conserved. A second option offered the converse: a low fixed payment and a high marginal payment for energy conservation. The third option was intermediate. The options were designed so as to induce truthful revelation of type. That is, a company with cheap options for conservation should prefer to choose the high-marginal-payment option, and a company with few conservation prospects should prefer the high-fixed-payment option. Further, the schemes were designed with the intention that the utilities should not make extraordinary profits, and that cost-effective DSM measures should be supported. The spirit of the approach is similar to that used by insurance companies in offering customers different combinations of deductibles and premiums.

Designing the menu options is not a completely scientific, objective process. The idea is to roughly guess the range of situations faced by different utilities, and put parameters on them in a way which seemed likely to satisfy the truthful revelation principle. In practice, the menu was designed through a collaborative stakeholder process as described in section 0. Lewis (1997) describes it as follows:

> To create these options, regulators solicited data and information on all aspects of the conservation program, including the utilities' projected baseline, levels of energy production, the costs of managing and marketing DSM measures, the energy savings resulting from conservation investments, and the demand for DSM measures by utility consumers. The process for collecting and analyzing data was a collaborative one. Input from all stakeholders including the utilities, residential and industrial customers, and conservation and environmental interest parties, as well as regulatory staff, was solicited.

What is significant for present purposes is that it was possible for a diverse group of stakeholders to design a politically-acceptable menu choice scheme.

Application to GHG ER's: the principle

Here is an example of how a revelation mechanisms might be applied to baseline determination. Suppose that some class of projects – for instance, fuel-switching projects in a particular countries – includes two groups of participants. Group L has low baselines, group H has high baselines, and it is difficult for an outside observer to distinguish between them. There are two broad sets of explanation for the difference between L and H:

1. *Differences in capital cost and pollution valuation* As we have discussed, one reason for sites to differ in baselines is that they face different capital costs and pollution costs. Facilities with higher capital costs and lower valuation of pollution damages will

tend to have higher baselines. They will also tend to have higher marginal cost of emissions reductions, since a facility of this type reckons less gains from fuel savings and reduction of local pollutants such as SOx and particulates.

2. *X-inefficiency*. Alternatively, group H might, for internal organizational reasons, have a higher level of x-inefficiency and thus have a larger supply of relatively low-costs GHG abatement options. In this case, the marginal cost of reductions would be lower for group H.

Figure 4 shows the dilemma of uniform crediting strategies. The 'true' baselines are bH (high baseline group) and bL (low baseline group). Marginal costs of emission reduction, as a function of total emissions, are MCH and MCL for the two groups. The market price of ER's is P. At this price, if baselines were observed by the certifying agency, the facilities would reduce emissions to cH and cL respectively. Group L facilities receive revenues of P*(bL-cL) and incur total costs equivalent to the area under MCL between bL and cL [which we will designate TCL(cL)], and similarly for group H. Since baselines are not observed, however, a group L facility is tempted to claim that it really belongs to group H. This gains it additional revenues of P*(bH-bL), and results in the crediting of bH-bL invalid reductions.

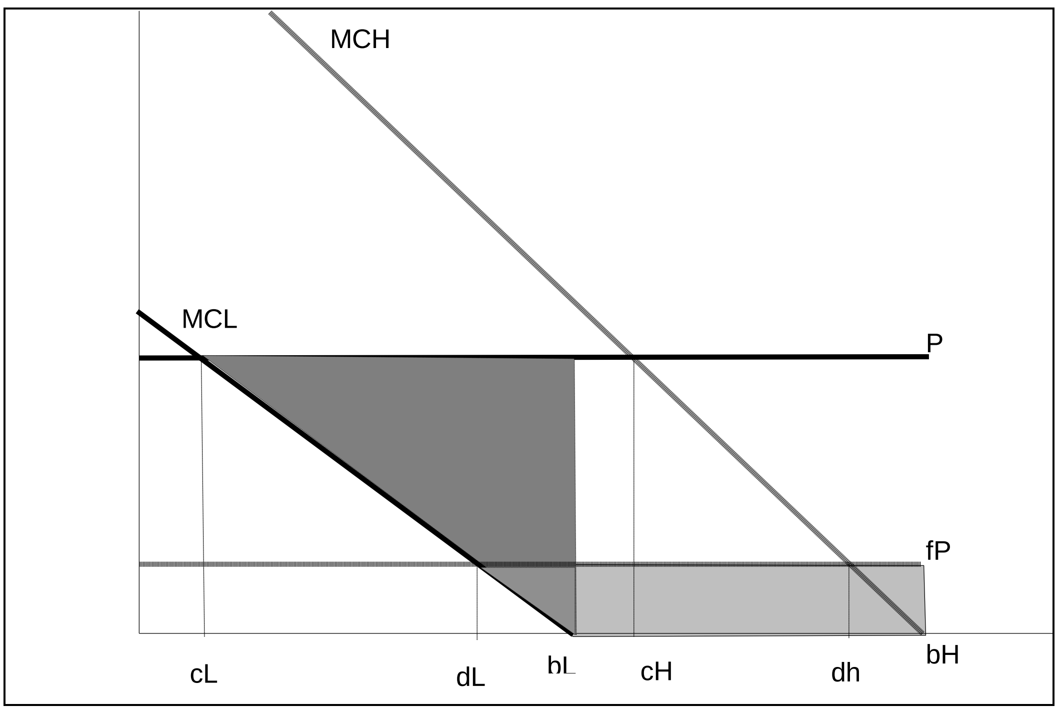

Figure 4

A partial crediting strategy is also unsatisfactory. Suppose that reported emissions reductions are credited at a fraction rate f<1, so that the effective price per reported emission reduction is fP. L's rents are reduced, and invalid reductions are reduced to f*(bH-dL). However, actual emissions are also reduced to (bL-dL) + (bH-dH), and H's rents are also reduced.

An alternative is to offer applicants for ER's a choice between two options:

- *option A:* receive full credit for emissions reductions below the stringent baseline bL
- *option B:* receive partial credit, at some rate f, for reductions below the generous baseline bH.

The revelation mechanism is successful in revealing truth if group L facilities choose option A and group H facilities choose option B. This in turn requires that:

- L's profits are higher under option A than under option B, so that:

$$P^*(bL-cL)-TCL(cL) > fP^*(bH-dL)-TCL(dL)$$

(the vertical striped area is greater than the horizontal striped area)

- H's profits are higher under option B;

(clear in the case illustrated because under this option H must spend TCH(bL) before earning any credits, and the face a marginal cost of reductions in excess of the price.

If an appropriate crediting fraction can be found, the revelation mechanism eliminates invalid ER's and reduces L's rents. Unlike the fixed-fraction partial crediting approach, it allows L to produce the efficient amount of ER's. However, relative to uniform full crediting, it screens out or reduces in scope some valid projects: once again, the type I vs type II error tradeoff. In this case, it denies H some credits.

Whether an appropriate menu can be found depends on the shape of marginal abatement curves and the degree to which they can be elucidated. An interesting real-world example of something akin to a menu mechanism was used in baseline definition for Costa Rica's Protected Areas Project (SGS 1997, Chomitz *et al.* 1998), the basis for an offering of CTO's (certified, tradable carbon offsets). The project defines what it regards as a defensible baseline, based on available information. It recognizes, however, that future information (such as improved studies of land cover change) may yield retrospective refinements in baseline definition. For that reason, it places about 20% of offsets from first-year activities in a 'permanent buffer'. In effect, the project only begins claiming offsets for sale after reductions exceed this buffered amount. This looks something like the outcome of a choice between partial crediting from a high baseline and full crediting from a low one, where the latter option has been chosen.

Application to GHG ER's: another example

A natural application of a partial crediting strategy of this sort occurs when there is agreement that additionality is uncertain. For instance, suppose that a firm has a number of possibilities for energy conservation, with payback periods ranging from six months to four years. The additionality of the six-month-payback measures is open to some question. Perhaps there are true barriers to adoption, perhaps the firm is inefficient and has set up its own, surmountable stumbling blocks. A Bayesian approach might suggest that partial credit be awarded to the six-month-payback measures, with increasing credit as the payback period lengthens.

Institutional aspects of menu-choice mechanisms

The process of determining the menu choices is involved, time-consuming and unavoidably political in nature; it cannot be reduced to an algorithm. For this reason, applications in regulatory economics typically involve direct negotiation between the regulatory principal and the entity being regulated. In the context of the UNFCCC, it might be difficult to delegate responsibility for these negotiations to a third-party certifying organization. The use of this mechanism might therefore be reserved for very large projects (or classes of homogenous projects), with negotiations undertaken directly with a designated supervisory body (such as that associated with the Clean Development Mechanism). For instance, it might be applied to the determination of baselines for national-level carbon sequestration projects, or national-level emissions related to electricity generation.

On the other hand, we should not entirely rule out the possibility that menu-choice schemes could be designed in a decentralized system. A collaborative process among stakeholders, described in more detail in section 0, might be empowered with designing menu-choice schemes.

The political acceptability of this kind of mechanism might be related to the explanation of the differences between high and low baseline scenarios. The menu-choice mechanism is more appealing when baseline differentials are thought of as being related to x-inefficiency, as in the example of differing payback periods for energy conservation schemes. Here the high baseline is somewhat questionable, justifying the application of partial credit. A firm claiming a low baseline, on the other hand, is one that has already undertaken extensive efficiency-increasing measures, and therefore deserve full credit for further reductions.

6. SYSTEMS BOUNDARY ISSUES

SYSTEMS BOUNDARIES: SPATIAL

Emission-reduction projects may have a variety of indirect effects, both positive and negative, on emissions elsewhere. Consider, for instance, a project which reduces the demand for coal by an industrial consumer. In a competitive market for coal, the result will be an infinitesimal decline in the overall market price, as the demand curve shifts slightly to the left. This infinitesimal decline, however, affects a very large number of other consumers, who compensate by increasing their consumption. Individually this response is negligible, but the collective response of many thousands of consumers will be a "snapback" increase in energy consumption and emissions which substantially dilutes the initial reduction.

A simple algebraic illustration: suppose that emissions are proportional to the consumption of an energy commodity, whose demand and supply curves are:

$$\ln Q_D = \ln d_0 + d_1 \ln P$$

$$\ln Q_S = \ln s_0 + s_1 P$$

$$d_1 < 0,\ s_1 > 0$$

A demand side management project reduces d_0 by an very small fraction. If there were no "snapback" effect, the proportional change in Q would be the same, that is the elasticity of Q with respect to d_0 would be 1. In fact, the elasticity is:

$$\partial \ln Q / \partial \ln d_0 = s_1/(s_1 - d_1)$$

where s_1 and d_1 are the price elasticities of supply and demand. Thus, quite intuitively, "snapback" disappears only where demand is completely price-inelastic, or where supply is perfectly elastic. As overall demand becomes more and more elastic, a larger proportion of the direct, observed emissions savings are vitiated by induced increases elsewhere in the economy.

In the case of exhaustible fossil fuel resources, changes in expected demand should lead to changes in the entire future time-path of depletion.[8] While it is difficult to predict how much of the rebound is contemporaneous, IEA (1995) describes the world coal market as being responsive in output and price to short-term demand pressures. It cites Australian econometric studies showing a short-run price elasticity of supply of about 0.4 and a long run elasticity of 1.9.

Market spillovers such as this abound in JI/ER projects. Consider the following examples:

1. *New private powerplants.* As electricity markets become deregulated, it becomes more difficult to predict the impact on emissions of a new, marginal plant. A new geothermal powerplant, for instance, may end up expanding the supply of electricity relative to the reference case, rather than displacing existing production, kilowatt-hour for kilowatt-hour. Net impacts then depend on how the plant's installation affects

[8] I am grateful to Sam Fankhauser and Luis Constantino for making this point. See also Fearnside (1997).

either the market price of electricity, or, if prices are fixed, how people are rationed into the system. These impacts also depend on the daily and annual timing of the new output.

2. *Forestry projects* A project which protects a plot of forest from conversion to agriculture may simply raise the demand for forest conversion elsewhere, by slightly raising the price of cattle, maize, or timber, and slightly reducing the reservation price of labor for forest clearance. In the limit, if demand for timber or for agricultural conversion is inelastic, protecting a plot of forest simply diverts conversion elsewhere, possibly even outside the national boundaries of the host country. (Brown *et al* 1997). On the other hand, establishment of new plantations may absorb labor and reduce the price of wood products, reducing pressures for deforestation.

3. *Coal efficiency projects*. Martin (1998), discussing within-facility reactions to changes in effective energy prices, emphasizes that increases in coal efficiency have the perverse effect of inducing switches from gas to coal, vitiating the direct emissions reductions. Michaelowa (1997) makes the same point. At the market scale, this may be a particular problem for projects which support coal-washing, which reduces coal transport costs and increases combustion efficiency. If successful, these may induce consumers to switch into coal from less emissions-intensive fuels.

4. *Industrial efficiency*. Any improvement in industrial efficiency may result in a decrease in price and increase in production for the good in question, along with a concomitant increase in emissions.

5. *Demonstration effects.* On the other hand, positive spillovers are also possible. A demand-side management project may have a demonstration effect, spurring the adoption of emissions-reducing technologies at other sites.

In all these cases, site-specific assessments of carbon emissions are inadequate to capture the scope of project impacts. It is necessary to take a wider, sectoral view of both baseline and project-related emissions. This can be done with varying degrees of sophistication.

The simplest approach is to use crude estimates of sector-wide supply and demand elasticities to estimate leakage effects. Martin (1998) provides simple examples of how this might be done. This approach ignores general equilibrium effects, which will be difficult to model in a practical fashion. In other cases, such as the electricity sector, relevant sector-wide models may be available "off the shelf".

Brown *et al*. (1997) stress that proper project design can reduce or eliminate these market-based leakages. For instance, a pasture abandonment project might arrange for the intensification of beef production in existing pastures so as to neutralize any market-mediated rise in the demand for pasture conversion. Since pastures to be abandoned will have very low stocking rates, this need not be expensive or difficult.

More generally, proper accounting for leakages will shift project selection efforts towards projects without these problems. For instance, a reduced-impact logging projects that maintains log output (relative to the baseline), but reduces collateral damage to non-harvested trees will not have any market-based leakages.

PERMANENCE AND CARBON SEQUESTRATION

ER projects have value because they affect the time-path of atmospheric GHG concentrations. Emissions abatement projects have different durations of impact than do sequestration or forestry projects. The difference needs to be explicitly accounted for when assessing baselines and calculating offsets.

It is easiest to illustrate the importance of *duration* in ER baselines through a parable. A buyer of carbon offsets is willing to pay $20/ton, the going price. His first purchase is from the owner of an abandoned coal mine. There is a ton of carbon on fire in the mine; if nothing is done, it will all burn instantly. For $20, the mine owner offers to put the fire out and ensure that it will never restart. This keeps the ton of carbon out of the atmosphere forever, and there are no market-mediated spillover effects. The buyer thinks that this certainly constitutes a one-ton offset, and buys it for $20. The buyer now walks down a road populated by farmers. Each farmer owns a tree containing one ton of carbon, and wants to slash and burn it in order to plant some crops. The first farmer is just about to chop down her tree and burn it when the buyer arrives. For $20, she offers to postpone cutting the tree for twenty years. She makes no promise about what will happen thereafter; since the tree is already worth cutting, and crop prices are rising, we presume that it will be cut as soon as the 20-year contract expires. The buyer agrees to the terms, and purchases a 20-year offset for $20. The second farmer will postpone cutting for only 10 years, but still demands $20. The buyer assents. The third farmer agrees only to postpone cutting for a single day, but the buyer, with some misgivings, agrees to this, too, and pays $20 for the third offset. A ton is a ton, isn't it?

This example is by no means artificial. Imagine a project that sells offsets that arise from secondary regrowth on abandoned pasture. Each ton of biomass accumulation is offered as a homogenous commodity, irrespective of when that ton is removed from the atmosphere and embodied as biomass. Suppose the commitment to the offset buyer is for a fixed period of twenty years. Thus the first ton of growth is guaranteed to be out of the atmosphere for nearly twenty years. The last ton is guaranteed to be sequestered for only days. The argument is the same for a fixed-duration deforestation prevention project. If the baseline is a constantly decreasing biomass over time, the last averted ton has a very short guarantee attached to it.

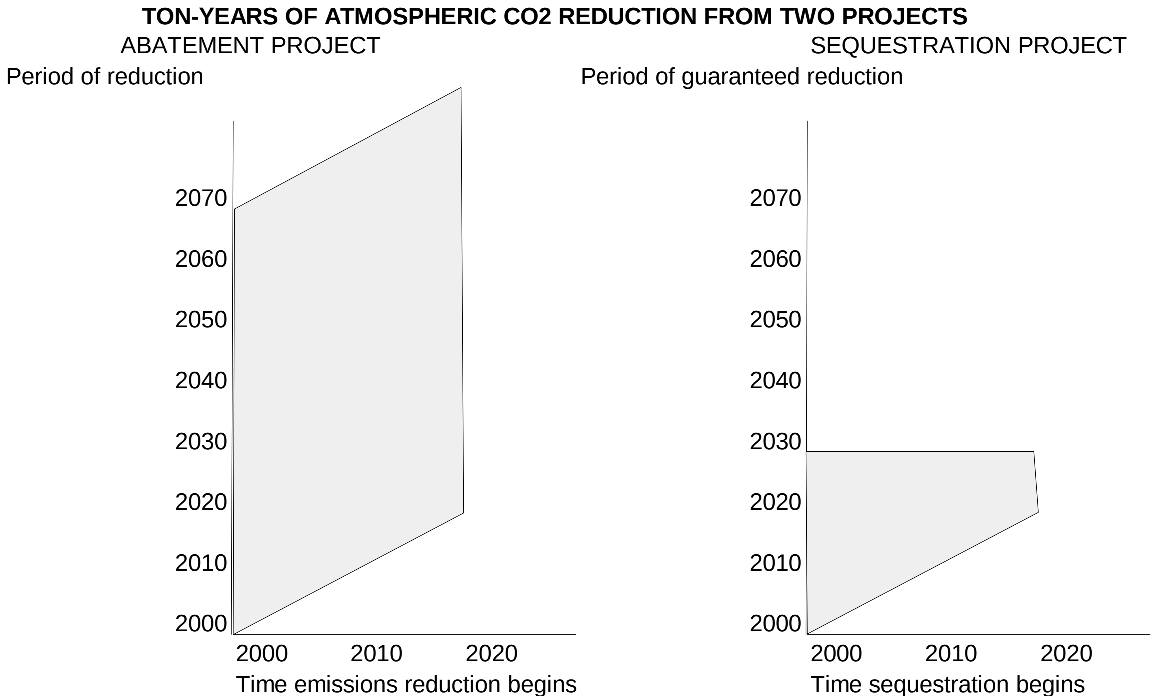

Figure 5

The issue is shown graphically in Figure 5. ER projects reduce atmospheric GHG relative to a baseline. This reduction in atmospheric carbon has both a start date and a notional end date. The horizontal axis plots the time at which an emissions reduction starts. For each ton reduced, the vertical axis shows the period during which this reduction affects atmospheric concentration of GHGs. In the case of an abatement project (left panel), the reference scenario envisions a flow of carbon being combusted at a constant rate from 2000 to 2020. Each ton of carbon is resident in the atmosphere for 70 years[9]. The first ton would have been combusted in 2000 and resided in the atmosphere until 2070. The last ton would have been combusted in 2020 and resided until 2090. The ER project completely abates these emissions. The shaded area shows both the quantity of resulting carbon offsets and their duration.

The right panel shows a project which sequesters carbon through forest regeneration. In 2000, biomass begins accumulating as secondary growth. The project contract concludes in 2030, after which landowners are free to burn the accumulated growth. To be entirely parallel to the abatement project, the sequestration project would have to provide a longer guarantee.

It would be possible to make sequestration and abatement projects commensurable by computing the number of discounted ton-years of atmospheric GHG reduction due to the project. This has been suggested by Fearnside (1997). A justification for this can be derived from Rosebrock's (1994) account of an optimal control model by Richards (1993).

[9] This assumption is for expository purposes only. The half life of CO2 is probably closer to 100 years, so that a substantial amount of CO2 is resident for centuries.

The optimal control model maximizes social welfare allowing for changes in atmospheric GHG concentrations and their effects on the economy. If economic damages are linear in gas concentrations, then the present value of a marginal reduction in gas concentration in year t is given by:

$$\lambda = \exp(-rt)*\gamma/(r+\delta)$$

where γ is a gas-specific constant, r is the social discount rate, and δ is the dispersion rate of the gas. By integrating this shadow value over time, we obtain the relative value of fixed-duration sequestration services relative to abatement or perpetual-duration services. (See table below)

	Years			
r	*1*	*10*	*20*	*40*
0.03	0.030	0.259	0.451	0.699
0.06	0.058	0.451	0.699	0.909
0.10	0.095	0.632	0.865	0.982

At a discount rate of 6%, a single ton-year of sequestration services is worth about 5.8% of the value of a perpetually sequestered ton; a twenty-year guarantee is worth 70%. Reckoning sequestration services in ton-years, rather than tons, solves an important problem hindering the development of markets for sequestration offsets: the credibility of long-term contracts. These pose an even more severe problem for sequestration projects than for abatement projects. For abatement projects, this year's and previous years' reductions are for all practical purposes perpetual. While there is some risk that next year's abatements may not be accomplished, the past year's achievements are secure and can be credited. For sequestration projects measured in undifferentiated tons, there is always the risk of reversal of achievements. If crediting depends on maintenance of a forest for a specified long period – say twenty years – there is always a risk that natural disaster or political upheavals towards the end of the period could undo the previous decades' accomplishments.

In a pay-as-you-go sequestration service scheme, a certifying authority would periodically check sequestered biomass against the baseline, and certify the accomplished number of ton-years of credit; ton-years could be aggregated into 'perpetual' tons at an established, fixed conversion rate. This device greatly reduces the risk to ER investors, and particularly facilitates agreements which involve ongoing payments for forest maintenance. It would allow the participation of countries or landowners with perceived high risks of nonperformance. These suppliers may have low costs of supply, but might be shut out of the market if only long-term sequestration contracts were valid.

Another advantage of ton-year crediting is alleviating concerns about loss of sovereignty. Some nations object to permanent or very long-term sequestration commitments, viewing them as equivalent to loss of sovereignty over their territory. But forestry investments for limited-term production of ton-years of emissions reductions are no more a loss of sovereignty than are investments in palm-oil plantations. Ton-year crediting allows host countries to determine a period of commitment with which they feel comfortable, and allows them to reclaim the stream of emissions reductions thereafter. This might smooth the long-term transition to an era when the host countries assume emissions limitations.

Ton-year crediting has already been applied, implicitly, to the analysis of reduced-impact logging projects. Sustained logging activities generate sawtooth-shaped graphs of carbon storage over time. (See figure) Carbon storage dips sharply when the stand is harvested, recovers with regrowth, and then dips at the next harvest. Reduced impact logging projects generate offsets by reducing the amount of collateral damage that loggers do as they extract marketable timber. For instance, reduced impact logging techniques minimize the area cleared for skid paths, roads, and landing pads, and use vine-cutting and directional felling techniques to reduce the number of trees unintentionally damaged in felling. This shifts the jagged time-profile of carbon upward. Boscolo *et al.* (1997) evaluate the GHG impact by computing the *discounted* carbon accumulation, which is equivalent to the discounted integral over time of the shaded area in the figure.

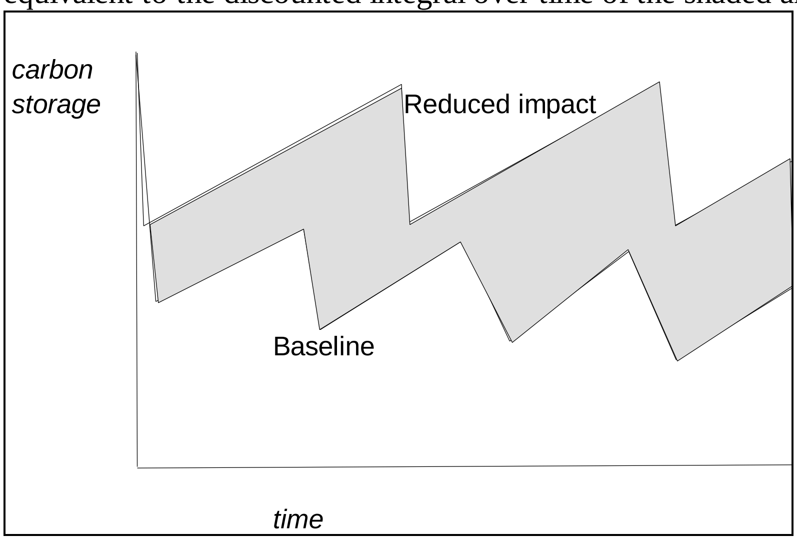

Figure 6 (adapted from Boscolo et al 1997)

7. DEMAND-SIDE-MANAGEMENT INCENTIVE SYSTEMS: LESSONS FOR GHG BASELINE METHODOLOGIES[10]

Demand-side management (DSM) incentive systems in US utilities constitute an existing, large-scale analog to the prospective GHG emissions reductions credit system. These programs reward utilities for saving energy (and as a byproduct, reducing GHG emissions) against a hypothetical baseline. A typical incentive payment is roughly equivalent to $20/ton of carbon[11]. What is interesting for our purposes is that these large-scale programs have had to grapple with all of the methodological problems involving baselines discussed above and have come up with practical, politically acceptable solutions.

This section describes DSM incentive systems, explains how they are analogous to GHG ER systems, and describes their methodological approach to establishing baselines.

DSM INCENTIVE SYSTEMS AS AN ANALOG TO GHG EMISSIONS REDUCTION CREDIT SYSTEMS

For about the past fifteen years, state regulators have charged utilities with providing energy services to customers at the lowest cost. There are two ways of doing this: through increased generation, and through DSM activities which increase efficiency of energy use. Typically, DSM programs involved providing incentives for customers to adopt higher-efficiency technologies for lighting, air conditioning, electric motors, and so forth. Thus the regulators had to find ways of rewarding the utilities both for producing and conserving electricity (Hagler-Bailly 1998). In order to reconcile these opposing objectives, it became necessary to specify a baseline against which energy savings are reckoned. Incentive payments to the utilities are then tied to these savings – sometimes in a simple linear fashion, sometimes through a complex formula.

The resulting analogy with GHG offsets is very close: the commodity, energy savings, is a kind of offset; the baseline is the energy consumption level against which savings are calculated and incentives credited; the state Public Utilities Commission is the analog of the UNFCCC, setting the rules of the market, including the definition and certification of baselines; the utility is the equivalent of the carbon credit buyer, and the customer is the seller. In fact, most DSM programs result in CO2 emissions reductions and therefore could potentially qualify as producers of GHG offsets.

---

[10] This section draws heavily from a background paper prepared by Hagler-Bailly Consulting, Inc.: "Evaluating Greenhouse Gas Mitigation through DSM Projects: Lessons Learned from DSM Evaluation in the United States". Verbatim quotes are double-indented.

[11] Eto (1995) reports mean shareholder incentives, weighted by program size, of 4 cents/kWh in a sample of 40 large commercial DSM programs; we apply a rough estimate of an emissions rate of 200 tons C/gWh.

In setting up these programs, the state Public Utilities Commissions have had to grapple with virtually all of the baseline issues discussed earlier. A key issue in energy conservation incentives is the extent of free-riders; i.e., whether the recipient of a conservation incentive would have undertaken the required conservation actions even in the absence of an incentive. Policy and implementation debates have also focused on "snapback" or "takeback" – the degree to which increased efficiency induces increased usage of energy, and spillovers, or the extent to which DSM activities have a demonstration effect and result in more rapid technological diffusion. Table 2 summarizes some of the parallels between DSM and GHG program issues and terminology.

Table 2 Related GHG Emissions Reductions and DSM Terminology

GHG Emissions Reduction Terminology	Related DSM Terminology
GHG emissions reduction credits or offsets	***Net program impacts.*** Energy use reductions that are attributable to the program, i.e., they would not have occurred had the DSM program not been offered.
UNFCCC Conference of Parties or subsidiary body such as the Clean Development Mechanism: regulator of credit creation and trade	***State Public Utility Commissions*** which set the incentive terms for utilities
host countries: producers of offsets	***utility customers:*** ultimate producers of energy reductions via use of energy-saving technologies
***investors, Annex I countries*:** finance activities in host countries which lead to production and acquisition of offsets	**utilities** provide incentives for customers to reduce demand, and receive credit
Additionality. Projects must provide for a reduction in emissions that is additional to any that would otherwise occur. (FCCC/CP/1997/L.7/Add.1)	***Free riders.*** Net impacts are often calculated by subtracting out free riders, i.e., those customers who would have installed the energy efficiency measure on their own without the DSM program.
Positive leakages. Degree to which project activities result in decreased emissions outside the project site or boundaries	***Spillover; Market Transformation; free drivers.*** Program spillover occurs when the program influences the program participants, and customers who do not participate in the program to invest in energy efficient technologies. Market transformation refers to spillover that permanently affects demand and supply of efficient technologies.
Negative leakages. Degree to which project activities result in compensating increases in emissions outside project site or boundaries	***Takeback.*** Takeback (also called snapback or rebound) is an economic response increased energy efficiency or lower effective energy prices.. For example, a homeowner may leave lights on longer after installing efficient light bulbs.
Static vs. dynamic baselines In the absence of the project, should emissions be assumed to have remained constant over time?	***Persistence*** of benefits: the extent to which net impacts should be adjusted over time due to changes in baseline or reference technology, as well as for actual performance of installed efficiency measures. ***Statistical adjustment*** of gross (before vs. after) energy consumption for nonprogram factors such as weather.

Source: modified from Hagler-Bailly 1998

SCALE OF DSM PROGRAMS

DSM programs are large and enjoy widespread support. Hagler-Bailly (1998) reports:

> From 1989 to 1996 about $16 billion was spent on these programs by regulated utilities. (EIA, 1996, 1997). In 1996, 1,003 utilities in the United States reported having a DSM program; these utilities represent almost one-third of all U.S. electric utilities, and include 97% of large utilities with generation levels of 120,000 MWh or higher (EIA, 1997).
>
> According to the EIA (1997), energy savings for the 573 large utilities totaled almost 62 billion kWh for 1996, which represents 2% of annual electric sales, at a mean cost of less than 3.0 cents per kWh saved. Commercial programs accounted for 47% of these energy savings (29 billion kWh), residential programs for 33% (21 billion kWh), and industrial programs for 17% (10 billion kWh).

HOW DSM PROGRAMS EVALUATE BASELINES AND ADDITIONALITY

Baselines are an implicit rather than explicit part of DSM program evaluation. Typically, net program impact is computed in two steps. *Gross impacts* are the difference between observed energy consumption by program participants (technology adopters), and imputed consumption, had the participants not adopted the technology in question:

gross impact for adopters=
predicted consumption assuming nonadoption-
observed consumption with adoption

Net impacts are gross impacts, adjusted for free ridership:

net impact for adopters=
gross impact*[1-probability of spontaneous adoption]

In principle, net impacts should also credit program-induced adoption by *nonparticipants* (spillover effects), though this is rarely done in practice.

Baseline determination enters the evaluation in both steps: in the adjustment of predicted consumption for extraneous factors, and the prediction of spontaneous adoption.

Gross impact evaluation

Methodologies for gross impact evaluation differ greatly in sophistication (Hagler Bailly 1988):

- *engineering methods* typically apply standard energy consumption factors to with and without-project technologies. They may for instance, use default or spot-metered energy consumption rates for old versus new appliances, multiplying by reported or assumed hours of operation. In some cases, however, sophisticated process models are constructed – for instance, to allow for the cross-effects of lighting choices on energy used for heating.

- *statistical methods* include a wide range of analyses of billing or other actual consumption data. They derive the baseline from observed utilization rates – either from pre-program data for participants, or from data on a comparison group. More sophisticated approaches statistically adjust the observed data for compositional differences in program and comparison groups, or for extraneous factors such as weather. At the limit, the distinction between gross and net impact evaluation disappears as econometric approaches are be used to model the program participation decision.

- *integrated methods* combine parameters derived from both statistical and engineering methods.

Table 3 shows the range of methodologies applied to different project types. Hagler-Bailly (1998) reports on choice of methodology:

> Method selection has depended on a number of factors including precision requirements, evaluation budget, and evaluator preferences or skills. CADMAC (1994a) provides detailed comparisons across impact evaluation methods of data demands, errors, cost, and robustness.
>
> Engineering methods alone can be inexpensive and simple to implement, and are appropriate for small programs or programs such as industrial motors or new construction that lack comparison groups or pre-program data. They also allow analysis of interactions between measures (e.g., efficient lighting can reduce demand for air conditioning in commercial buildings), and analysis of load profiles and time differentiated impacts (e.g., peak versus off-peak impacts[12]) (RCG/Hagler Bailly, 1991).
>
> However, engineering methods as a stand-alone approach are now rarely used. Most engineering approaches take advantage of the power of sampling and statistics to generate a new set of engineering-based methods such as CEM or SAE methods. A more complete discussion of these methods is found in CADMAC (1994a).
>
> ...As an example of the variety of methods used in practice to estimate energy savings, a study of 40 of the largest commercial lighting programs (Eto et al., 1995) found a wide variety of evaluation methods, including instances of

[12]. This characteristic is important for estimating GHG reductions because generation fuel and emissions rates often differ between peak and nonpeak power sources.

> programs using multiple evaluation methods.... Thirty five utilities... verified measure installation through on-site inspections, and 20 verified hours of operation through on-site inspections. Nineteen utilities, or almost 50%, also conducted billing analyses: one used a simple pre- and post-program comparison for participants; three used simple billing comparisons including nonparticipants; three used regression methods; and 12 used SAE regression methods.

In general, significant progress has been made in gross impact evaluation, with increasingly well-defined protocols for measurement and analysis of energy consumption data. Much of this work would be directly applicable to JI/ER projects.

Net impacts

The methodology for assessing net impacts is less well developed than that for gross impacts. DSM evaluations use all three methods described in section 3: direct questioning of participants, control groups, and behavioral models. Marbek Resource Consultants and RCG/Hagler Bailly (MRC and RCG 1994) provides a concise summary of baseline and free-ridership methodologies as applied to standard types of programs (see Table 4).

In general, the *direct survey* approach has been the most widely used. It is usually directed towards measurement of free ridership, though occasionally used to assess spillover effects. *Comparison groups* are often used to adjust for weather and other factors affecting energy consumption. Sometimes informal, rather than statistical comparisons are made of building practices or appliances. (Hagler Bailly 1998). More rarely, *behavioral models* are constructed along the lines sketched in section 4. These can be important, for instance, in industrial motor replacement programs where the reference scenario can be continued use of the old motor, rewinding, or replacement.

Of these techniques, the comparison group approach seems the most transferable to GHG ER programs. Direct survey approaches may not be reliable. Baseline scenario definition for new construction is extremely problematic, since buildings are idiosyncratic systems of energy-using components. It is possible to use engineering methods to simulate the energy consumption of a candidate reference scenario (see DOE 1997) but justifying that scenario is difficult.

Table 3. Summary of Impact Evaluation Techniques

Technique		Appropriate Program Types	Relative Cost Requirements**	Relative Precision	Comm
Engineering	Algorithms	Lighting (residential and commercial), water heating, refrigeration, motors, some processes	Low	Low/medium	Precision can be increased thro nonprogram impacts and calibr
	Building simulation models	HVAC,* daylighting	Low/medium	Low/medium	Precision can be increased thro nonprogram impacts and calibr
	Detailed process/application-specific models	Thermal cool storage, industrial processes	Medium	Medium/high	Models often rely on end-use n expensive and is used when sav large.
Statistical	Simple comparison	Residential HVAC, water heating, small commercial HVAC, lighting	Low	Low	Comparison groups must be cl variation is not addressed.
	Augmented comparison	Residential HVAC, small commercial HVAC, some industrial processes	Low/medium	Medium	Comparison groups must be cl comparison may be useful for i
	Multivariate regression	Residential HVAC, water heating, commercial HVAC, lighting	Medium/high	Medium/high	Typically requires data on both nonparticipants along with a su explanatory variables for the m
	Multivariate regression with participation model	Residential HVAC, water heating, commercial HVAC, lighting	Medium/high	Medium/high	Requirements are similar to mu sample size must be larger to a participation model.
Integrated Engineering and Statistical	Statistical audit procedures	Residential HVAC, water heating, commercial HVAC, lighting	Medium	Medium	Generally uses small samples o verify the engineering estimates
	Statistically adjusted engineering models	Residential HVAC, water heating, commercial HVAC, lighting	High	High	Precision will increase and cost engineering estimates are comp

* HVAC is heating, ventilation, and air conditioning.
** Although costs depend on several factors such as program size, the approximate categories indicated are low (<$100,000), medium ($10
high (>$500,000).

Source: Hagler Bailly (1998) based on Marbek Resource Consultants and RCG/Hagler, Bailly (1994).

Table 4. Summary of Baseline Issues for Selected Energy Efficiency Measures

Measure	Baseline Approach	Free Rider	Spillover	Takeback	
Residential New Construction	Statistical analysis (comparison area), builder survey inputs for engineering model	Builder survey, sales data, or survey in comparison area	Time series comparison of building practices, builder survey, survey in comparison area	Statistical bill comparisons or home-buyer survey	Persistence not a sign DSM measures
Comments: Engineering analysis for small/informational programs; multivariate regression or SAE recommended for most incentive programs; recommended for large programs.					
Residential Envelope	Statistical analysis (pre/post or comparison groups), engineering analysis with on-site surveys	Survey (participants or both participants and nonparticipants)	Similar to residential new construction	Survey questions could address takeback issues	Not important for rela (insulation and windo caulking and weathers
Comments: Building simulations would be appropriate for small or informational programs, and statistical or integration methods are more app and incentive programs. Load programs, which require participants to complete much paperwork, are more susceptible than low-income weath					
Residential Hot Water	Engineering estimates	Survey (participants or both participants and nonparticipants)	Survey (participants or both participants and nonparticipants)	Survey (studies do not show large takeback incidence)	Water heater wrap an persistent (particularly than efficient heaters
Comment: Engineering analyses with surveys are recommended for small or informational programs. Statistical methods may not be able to eff and aerator impacts, but can identify impacts of hot water packages.					
Residential Refrigerators	Time series billing and metering data can be used for buy-back programs, but not new dwelling installations; engineering analyses can use standard efficiency levels as baselines	Pre-program sales (e.g., dealer survey) or participant surveys for incentive programs, participant surveys for buy-back programs that purchase second refrigerators	Survey (dealers, participants, or both participants and nonparticipants)	Participant surveys can ask about whether incentive led to different purchase, or if new refrigerator was purchased earlier than planned and old refrigerator kept as secondary refrigerator	Persistence is likely fo for buy-back program
Comments: Statistical analysis for incentive program will lack pre-participation data on alternative refrigerator purchase, but engineering metho for comparable, nonqualifying models. Buy-back programs are amenable to statistical methods because savings are potentially large and pre-pr					
Source: Hagler-Bailly 1998					

Table 4 (cont.)
Summary of Baseline Issues for Selected Energy Efficiency Measures

Measure	Baseline Approach	Free Rider	Spillover	Takeback	
Commercial New Construction	Engineering methods combined with surveys of nonparticipants/trade allies (efficiency levels), and participants (hours of use); comparison group selection is difficult because of structural heterogeneity	Participant surveys	Surveys of nonparticipating builders and designers	Not likely to be a significant factor	Relatively pers rates/remodel
Comments: Larger programs may be able to eventually use statistical or integration methods (e.g., SAE) to estimate baselines because the will be greater. Engineering methods are more common, e.g., building simulations (HVAC and daylight measures) or algorithms (e.g., ligh be improved by on-site monitoring or metering and statistical sample selection..					
Commercial Lighting	Engineering methods with survey or site-based data on baseline technologies and usage levels; statistical or combination methods may be justified by large expected savings	Can be large for lighting programs, use participant surveys; comparison groups can be difficult to identify	Survey (participant, or participant and nonparticipant) and equipment dealer surveys	Not likely to be a significant factor	Malfunctioning and remodeling
Comments: Engineering methods are commonly used for lighting programs, with SAE methods used for larger programs.					
Industrial Motors	Engineering methods or time series comparisons are commonly used	Can be large, especially among large participants, use participant and dealer surveys	Can be large, survey nonparticipants or estimate market saturation levels	Not likely to be a significant factor	Generally pers
Comments: If greater precision is required, time-series comparison approaches of end-use metered data are most appropriate. Identifying b motor, rewind existing motor, continued use of existing motor) is important.					
Sources: Hagler Bailly (1998) based on Marbek Resource Consultants and RCG/Hagler, Bailly (1994).					

Incidence of "free ridership"

Hagler Bailly (1998) summarizes metaanalyses of free ridership as follows:

> Saxonis (1995) reviewed the treatment of free ridership behavior in about 100 program evaluations. ...the diversity of approaches to estimating free ridership behavior (e.g., self-reported, or based on billing analysis or life-cycle analysis methods) continues to generate a wide range of free ridership incidence, e.g., 0% to 42% in a sample of 25 program evaluations of residential compact fluorescent bulb (CFB) programs, and 0-73% for a sample of 20 commercial lighting programs. However, when outliers were excluded from the sample, free ridership levels were less than 20% (and a majority less than 10%) for residential CFB programs, and less than 25% for commercial and industrial lighting programs.
>
> Comparable results are found in Eto et al. (1995), which analyzed a cross section of 40 commercial-sector DSM programs. Reported incidences of free ridership among the programs ranged from 0% to 50%, with a simple average of 12.2% and a standard deviation of 11.4%.
>
> Both utilities and the regulators have been interested in increasing net program benefits by reducing free ridership. One approach has been to analyze the correlates of free ridership, and then more aggressively market those customer segments least likely to adopt DSM in the absence of the program. A second approach has been to shift away from financial rebates towards information diffusion programs. While free ridership has never been eliminated (except perhaps for low income programs), it is possible to keep free ridership in programs such as lighting and motors to less than 30%.

Variations in free ridership measures are driven by variations both in programs and in the reliability of measurement methods. Practitioners in the evaluation industry believe that any underestimates of free-ridership are counterbalanced by the lack of credit for positive spillover effects.

Persistence and dynamic issues

DSM incentive programs routinely employ baselines that are adjusted after the start of the project. As the preceding discussion makes clear, baselines are often determined from control group data collected during the course of the DSM program. These data allow implicit or explicit adjustment for weather or other unpredictable factors affecting energy consumption. In addition, estimates of free ridership are often based on surveys conducted well after program initiation. This is an interesting counterpoint to the standard JI or GHG practice of pre-establishing a baseline.

DSM practitioners have been concerned with assessing the "persistence" of DSM actions: for how long after installation should a technology be credited with providing energy savings? This is primarily viewed as a monitoring, rather than a baseline issue. Energy savings from a high-efficiency light or motor may accrue over many years, but cease if the device is removed, unutilized, or broken. To ensure accurate reckoning of savings over time, it is necessary to check that the installed device (such as a high-efficiency light or motor) is still in place and functioning correctly.

In practice, DSM programs often employ periodic inspections to ensure that past installations are still functioning. Hagler Bailly (1998) report:

> The California protocols specify periodic persistence evaluations and full load impact evaluations, and the schedules differ across program types because expected measure lifetimes differ (CPUC, 1993). Retention studies, which determine whether efficient measures are in place, are usually required biennially after the installation year, and load impact studies are required three or four years after the installation year. These inspections need not continue indefinitely, however. (*see the discussion of the EPA Verification Protocols in section 6).*

a) In principle, the baseline can also change over time, as technologies, regulations, prices, or capacity utilization change. In practice, baselines are sometimes retrospectively adjusted to reflect changes over time in equipment utilization rates. This information is generated on an ongoing basis as utilities determine baselines for new participants. However, it is rare to retrospectively revise a baseline assumption about technology choice for prior-year participants (e.g. to have a dynamic baseline for the retirement date of existing equipment).

Evaluation costs

NARUC (1994) surveyed twelve states and found that utilities spent from 3% to 10% of their DSM program costs on evaluation, with a mean of 6%. The reviewers suggested that 4% to 8% was a reasonable guideline. EIA (1995) surveyed the 50 utilities reporting the largest amount of energy savings. Among 20 respondents, the mean proportion of DSM program costs devoted to evaluation was about 3%, with the maximum under 7%. Eto *et al.* (1995) also find mean evaluation costs of about 3% in a sample of 37 reporting large utilities in 1992.

There are almost certainly significant economies of scale in program evaluation. This is particularly true for statistically based estimates, since the sample size necessary for a given level of accuracy is more or less independent of the size of the population from which it is drawn.

Guidelines and protocols

To what extent can baseline procedures be standardized? Interestingly, there is no industry-wide standard for baseline determination. Although most states have some form of DSM incentive program, official protocols for energy savings evaluation have been established only by California and New Jersey. (Hagler Bailly 1998). Protocols for energy monitoring and verification have also been published by the EPA (1995) and by the US Department of Energy (the International Performance Measurement and Verification Protocol, USDOE 1997). State-of-practice reviews include EPRI (1995), EPRI (1996) and CADMAC (1996).

These guides and protocols contain, in many cases, quite detailed specifications for the metering and sampling of realized energy consumption. They may for instance specify minimum sample sizes, frequency of inspections, and metering methods. They are much less prescriptive with regard to the determination of net, rather than gross impacts – i.e., the determination of additionality. The California Protocols (CPUC 1992) merely suggest that net impacts can be assessed by comparing before-and-after consumption by participants with before-and-after consumption by a comparison group. The IPMVP simply takes pre-installation consumption as the baseline for retrofit/replacement programs (DOE 1997).

DO DSM PROGRAMS YIELD GENUINE SAVINGS?

As we have seen, DSM incentive programs face the same problems of offset definition, monitoring, and verification as GHG ER programs. The DSM programs have faced this challenge with increasingly sophisticated evaluation programs. Particular progress has been made in the area of monitoring actual energy usage and adjusting it for exogenous determinants such as weather. However, adjustments for free ridership (additionality) have been criticized by some as naive, crude, or methodologically flawed (Joskow and Marron 1992; Train 1994). On the other hand, spillover benefits are rarely measured. On net, are these programs actually generating additional energy savings?

Parfomak and Lave (1996) use a clever approach to test the aggregate accuracy of reported savings from energy conservation programs. Using panel data on 39 utilities for the period 1970-1993, they regress the annual change in electric sales on changes in electricity price, fuel prices, manufacturing employment, non-manufacturing employment, heating degree days, cooling degree days, and the reported net addition to conservation. They argue that a coefficient of -1 on the conservation report would show it to be accurate; a coefficient of 0 would show it to be spurious. Their estimated coefficient is -0.994, with a standard error of .281, strongly supporting the hypothesis that reported energy savings are meaningful. There is however a slight qualification: this result is obtained with the inclusion of a separate variable representing reported conservation by Southern California Edison, the largest utility in the sample and the reporter of the largest level of conservation. The SCE coefficient was estimated at -0.261, with a large standard error of 0.452 – not significantly different from either 0 or -1. They note that savings from SCE's purely informational programs (as opposed to those offering customer incentives) are widely regarded as overreported. However, this class of savings does not earn the utility itself an incentive award from the Public Utilities Commission and so is not regarded with concern.

While more meta-evaluation data of this kind would be useful, it appears that DSM incentive schemes, for all their potential shortcomings, have been successful. They have been widely adopted, indicating broad political support. The next section examines the institutions that support the integrity and credibility of these systems.

WHAT KEEPS BASELINES HONEST?

As the preceding discussion makes clear, program evaluation is not a pure science: a variety of methodologies are used, and adjustments for additionality (free ridership) are sometimes crude. We would expect that both financial incentives and evaluators' professional enthusiasm for DSM would tend to result in overstated energy savings. (This doesn't necessarily presuppose unethical behavior, just a persistent tendency to err in favor of higher energy savings when there is legitimate uncertainty.) Yet the results cited above suggest that on average, baselining is reasonable accurate. Does reliance on third-party evaluation suffice to keep baselines honest?

A cautionary example: transit forecasting

In "A Desire Named Streetcar" Pickrell (1992) presents a cautionary lesson from another interesting analog to baseline determination: the financial analysis of large public transit projects. Over the past three decades, large US cities have chosen among competing plans for public transportation. Because of the magnitude of the federal subsidies involved – over $60 billion – the federal government required cities to justify their choices among alternative projects on the basis of ridership and cost projections. As in the case of GHG baselines, these projections of hypothetical futures were the basis for the award of external funds to the cities. Generally these projections were undertaken by third-party consultants presumably concerned about their reputations.

Pickrell retrospectively analyzes the validity of ridership and cost projections in eight cities which chose to invest in rail transit, the most expensive and heavily subsidized of the transit options. He makes the following case for pervasive bias in the system:

- the decision in favor of a rail transit project was generally made on the basis of narrow projected advantages in ridership and cost
- transport demand modelling and cost projection can draw on well-established, increasingly sophisticated, methodologies; nonetheless
- in seven out of eight cases, projected ridership was less than half of actual ridership; in the eighth case, projected ridership was 28% below actual ridership;
- actual construction costs exceed projected costs by 17% to 150%;

- most of the projection errors are not traceable to errors in assumptions such as population growth rates, fares, and travel speeds.

Pickrell concludes that:

> ...By tolerating pervasive errors of the consistent direction and extreme magnitude documented here, the transit planning process has been reduced to a forum in which local officials use exaggerated forecasts to compete against their counterparts from other cities...Such competition increasingly leads officials to encourage their planning staffs and consultants to underestimate rail transit projects' costs and overestimate their prospective benefits. (Pickrell 1992, p 169)

DSM incentive systems and "collaborative process'

In contrast to the case of public transit, we have cited evidence from Parfomak and Lave (1996) suggesting that utilities' estimates of energy savings are reasonably accurate on average. Since utilities are rewarded for energy savings, and since the quantification of energy savings is usually performed by utility-hired consultants, this is striking. What maintains the system's accuracy?

Part of the answer is an active network of professional associations and meetings. (Hagler-Bailly 1998) A biennial conference series sponsored by the American Council for and Energy-Efficient Economy, the Department of Energy, and several utilities has produced reference literature and promoted networking. The Electric Power Research Institute has sponsored research and dissemination in evaluation methodology. And the Association of Energy Service Professionals' largest standing sub-committee is the one concerned with evaluation. These groups, publications, and activities promote consensus in evaluation standards.

Public oversight of the process is probably a crucial factor promoting accuracy in assessment. Although utilities generally prepare their own evaluations of energy savings, these evaluations must be reviewed and approved by the Public Utilities Commission (PUC). Hagler Bailly (1998) reports:

> This review process often includes outside parties as advisors and many states have established a formal "collaborative process" that includes environmental interest groups, ratepayer groups, industrial representatives, and others who might have an interest in the outcome of these DSM programs. A key component of this review is consideration of the evaluation plan....
>
> [The collaborative] process uses a review committee of interested parties established by the state PUC that reviews all DSM ... activities on a regular basis, e.g., quarterly or every six months. The utility files a report that presents the current status of each program in terms of its implementation

> in the field as well as on-going evaluation efforts. This regular communication between an oversight organization and the utility proved valuable to both parties. This allowed the utility to make real-time changes in their programs if problems were found, or to revise their evaluation methods if circumstances dictated that a different approach be used. These were presented to the collaborative committee and feedback was given to the utility.

In other words, ***baseline methodologies are subject to review not just by a hired consultant, but by an independent board of stakeholders.*** In some cases, this board reviews evaluations commissioned by the PUC. In the cases of Michigan, however, the evaluator reports not to the utility but to the committee itself, which consists of two representatives from ratepayer groups, and one each from the PUC, the utility, and the attorney general's office (NARUC 1994).

Conclusions on keeping baselines honest

The ER baseline determination process should not rely only on self-regulation by certifiers or consultants. Development of professional bodies can help to establish standards of practice. Development of an accreditation mechanism for certifiers is probably desirable, but may not be sufficient to ensure unbiased baseline setting.

Emulation of the DSM's "collaborative process" is an interesting complement to accreditation-of-certifiers. In the GHG context, it would involve inviting representatives of the public interest, probably from NGOs, to sit in on the baseline determination process. It is probably not even necessary to give these stakeholders a formal veto; rather, failure to achieve consensus might be noted in the course of the certification process, and might be expected to increase the likelihood that the certifier is audited by the accreditation board. The process need not, however be confrontational, and it is quite conceivable that stakeholders might decide that a proposed baseline was unreasonably strict. An advantage to the collaborative process is that it could be adopted voluntarily by project hosts and investors.

8. CONCLUSIONS AND RECOMMENDATIONS

To set up a baselining system, two levels of decision are necessary. First, some general guidelines need to be established for the determination of baselines and additionality. Arguably some of these guidelines need to be established at the level of the UNFCCC, but until binding rules are established, offsets producers and traders will need to make some provisional decisions. I outline the main issues and choices in setting guidelines. Second, task managers or project sponsors need guidance on selecting and applying appropriate baseline methodologies. A preliminary procedural sketch is provided.

SETTING GUIDELINES

Ground rules

a) *Should baselines be evaluated under prevailing prices and policies, or in a hypothetical distortion-free policy environment?*

Some, perhaps many, projects make sense only if policy distortions are taken as given. Alternatives include:

i) always compute baselines under distortion-free assumptions.
ii) decide, on a country-by-country and policy-by-policy basis, which policies are immutable in the short to medium run, for practical purposes. For instance, current energy subsidies could be accepted for baseline purposes if the host country has adopted a plan for their gradual phase-out.
iii) accept prevailing policies and prices for baseline purposes.

The most conservative option, given an emphasis on producing high-quality offsets, would be to invest only in projects which generated offsets under distortion-free assumptions. This is consistent with GEF guidelines.

There may eventually be a UNFCCC ruling on this issue. In the interim, project sponsors may wish to maintain two baselines, with and without policy distortions. They could reckon offsets against the more stringent baseline, but retain the option to use the higher baseline if officially permitted.

b) *(Relatedly) What assumptions should be made about the level and enforcement of regulations on air pollution (or alternatively, of the effective level of pollution charges or local environmental damages). Similarly, what assumptions should be made about the effective level of enforcement of forestry and land use regulations.*

Choices include:

i) Using charges or practices established under official regulations (even if unreasonably strict or lax).
ii) Adopting default values.
iii) Quantifying and adopting current effective practice.

Choice i) is the most straightforward to apply, but problematic for several reasons. Where standard are very low, there could be questions about moral hazard and fairness. Where standards are unrealistically high, and not enforced, their use would scuttle potentially valuable projects. Choice iii) is the closest to a true measure of additionality, if practice could be perfectly observed, but there are problems with measurement and again with moral hazard. Choice ii) is inevitably somewhat arbitrary, but could be based on either 'best practice' among comparable countries, or on a damage function estimate.

c) *What assumptions should be made about the effective cost or availability of capital for projects similar to those under consideration?*

Possibilities include periodic surveys of banks or businesses outside potential JI/ER sectors. It may be possible to come up with rough rules of thumb based on central bank lending rates plus a differential.

Independent baseline review

I have argued that it would be useful to have disinterested reviewers examine proposed baseline calculations, in addition to third-party certifiers. This has the potential advantage of significantly boosting the credibility of the baseline, but also introduces potential extra costs and delays. One way to proceed would be to incorporate the examiners into the baseline design process, rather than adding a final, time-consuming review step. This has the potential advantage of early identification of projects that may not be perceived as additional.

It may be worth experimenting on a trial basis to see whether this feature is worthwhile. Selection of neutral reviewers will be a key to success.

Project selection

It has sometimes been argued that baseline, additionality, leakage, and measurement problems are greater for DSM and forestry projects than for other classes, such as fuel-switching projects. This conventional wisdom could bear some reexamination:

- The clarity of fuel-switching baselines is illusory if there is a chance that the host would spontaneously switch from the reference technology to the proposed project.
- As in the case of DSM projects, some fuel-switching projects appear economically rational to adopt without external support. The existence of barriers to the adoption of profitable technologies, and of failure to account for local environmental benefits, is often less plausible in the case of large projects than for the assemblages of smaller activities which constitute DSM projects.
- Unlike DSM projects, there is generally no way to construct a credible control groups.
- Any project which reduces the demand for energy in general, and coal in particular, should be presumed to have significant leakage effects through market 'snapback', as reduced demand depresses prices, prompting nonproject consumers to increase their consumption.

This is not to argue against the fuel-switching projects, many of which are doubtless valuable, but rather to argue that other classes of projects deserve a closer look. There are for instance forestry projects which may have little or no leakage. These include:

- reductions in collateral damage by loggers to nonmarketed timber
- reductions in anthropogenic fires affecting nonmarketable timber
- projects which neutralize leakage. For instance, a pasture abandonment project might slightly decrease the supply of beef; but since stocking ratios on affected lands would likely be low, it would be possible to sponsor a compensating increase in beef supply from areas of intensive production so as to neutralize any tendency for pasture to expand elsewhere. (See Brown *et al* 1997 for other examples).

BASELINE DETERMINATION AT THE PROJECT LEVEL

Here is an outline of a step-by-step guide.

1. *First determine whether there is a natural comparison group for this project.* This may be the case if the project consists of a large set of small-scale industrial, residential, or farm interventions, and if there are similar 'control' units for observation outside of the project area. In some cases, it may be possible to construct comparison groups for individual large projects if they are suitably generic. If it is possible to construct a comparison group, the practitioner can draw on a broad set of methodologies developed for DSM and program evaluation.

2. *If comparison groups are infeasible, set up the baseline problem as an investment decision among several potential reference projects and the proposed ERC project.*

a. Describe the potential project choices. Use observed data or engineering/agronomic models to establish emissions rates, conditional on those choices.

b. Establish the values of key parameters: current and expected future fuel and electricity prices; pollution charges, shadow-prices, or regulations; capital costs or target rates of return. Ideally these should be set by default. In practice, early projects will establish precedents. This should be recognized in budgeting for project appraisals.

c. Propose an investment-choice decision procedure: how would the sponsor choose among these alternatives in the absence of PCF funding. Where the baseline is a Bank or IFC-financed project, apply standard evaluation tools: why would alternative A be financed and not proposed project B, in the absence of PCF funding? This reduces to incremental cost analysis in many cases. A very simple, spreadsheet-based model of NPV or IRR could be used at the pre-screening stage, with a more sophisticated model used for project appraisal.

d. Use this procedure to determine which alternative would be chosen in the absence of offsets funding. Confirm where possible by reference to current practice. (That is, even where a statistically valid control group cannot be constructed, it is important to know whether similar fuel-switching projects, new construction practices, etc. are being undertaken without subsidies.) If the predicted choice is ambiguous, consider a partial crediting strategy.

3. *Devise a protocol for measuring the actual emissions of the project technology, and the predicted emissions of the baseline or reference technology*. The former can be accomplished through monitoring and sampling procedures, and protocols for this purpose exist. The latter could be based on pre-project data, on a survey of comparison facilities, or on engineering models.

4. *Decide whether or not to use a dynamic baseline – one that is flexible over time*. Dynamic baselines will be worthwhile:

i) in replacement/retrofit projects, when retirement of the existing facility is sensitive to unpredictable changes in prices or interest rates
ii) when emissions are volatile because of variable and unpredictable facility loads

In the absence of control groups, dynamic baselines are constructed by i) modeling the retirement decision as in 2d) above, but with annually updated parameters; ii) using an engineering model to represent the response of the reference technology to alternative loads.

5. *Assess market and/or leakage impacts*. Market impacts are necessary to compute emissions reductions from projects that affect electricity supply or demand. Leakage impacts must be computed for most projects. For projects with marginal sectoral impacts, use default assumptions about market responses (such as 'snapback' responses to reductions in coal demand, or world timber market responses to project-sponsored reductions in log harvesting). For projects with nonmarginal sectoral impacts (such as the construction of large generating facilities), use sectoral models such as integrated resource planning models.

9. BIBLIOGRAPHY

Baron, David P. 1989. "Design of Regulatory Mechanisms and Institutions". In R. Schmalensee and R.D. Willig, eds., *Handbook of Industrial Organization, Volume II.* New York: Elsevier.

Boscolo, Marco, Joseph Buongiorno, and Theodore Panayotou. 1997. "Simulating Options For Carbon Sequestration Through Imporved Management Of A Lowland Tropical Forest". *Environment and Devleopment Economics 2:* 241-263.

Brown, Paige, Burce Cabarle, and Robert Livernash. 1997. *Carbon Counts: Estimating Climate Change Mitigation in Forestry Projects.* Washington: World Resources Institute.

Carter, Lisa, 1997. "Modalities for the Operationalization of Additionality". Prepared for presentation at the UNEP/German Federal Ministry of the Environment. International Workshop on Activities Implemented Jointly. Leipzig, Germany, March 5-6, 1997.

CADMAC. 1994a. "An Evaluation of Statistical and Engineering Models for Estimating Gross Energy Impacts." Prepared for California Demand Side Management Advisory Committee: Subcommittee on Modeling Standards for End Use Consumption and Load Impact Models, by R. Ridge, K. Parikh, D. Violette, and D. Dohrman. CADMAC Study I.D.: 2001M. Reissued by the Association of Energy Services Professionals, Boca Raton, FL, October 1996.

CADMAC. 1994b. "Quality Assurance Guidelines for Statistical and Engineering Models." Prepared for the California Demand Side Management Advisory Committee, by R. Ridge, D. Violette, and D. Dohrman. CADMAC Study I.D.: 2001M. Reissued by the Association of Energy Services Professionals, Boca Raton, FL, October 1996.

Chomitz, Kenneth M., Esteban Brenes, and Luis Constantino. "Financing Environmental Services: the Costa Rican Experience and its Implications".

CPUC. (California Public Utilities Commission) 1993. "Interim Opinion on Ex Post Measurement and Evaluation." California PUC Decision 93-05-063. May 19.

DeCanio, Stephen J. and William E. Watkins. 1998. "Investment in Energy Efficiency: Do the Characteristics of Firms Matter?" *The Review of Economics and Statistics 80:* 95-107.

GEF, 1996. *Incremental Costs.* Document GEF/C.7/Inf. 5, Feb 29, 1996.

Goldberg, M.L., and K. Scheuermann. 1997. "Gross and Net Savings Analysis for Unique Projects." In *The Future of Energy Markets: Evaluation in a Changing Environment*. Chicago, IL: National Energy Program Evaluation Conference, pp. 389-395.

EIA. 1997. "U.S. Electric Utility Demand-Side Management 1996." DOE/EIA-0589(96). Washington, D.C.: Energy Information Administration.

ELI (Environmental Law Institute) 1997a. *Tranparency and Responsiveness: Building a Participatory Process for Activities Implemented Jointly Under the Climate Change Convention.* Washington: ELI.

ELI (Environmental Law Institute) 1997b. "Establishing Carbon Trading Programs: Lessons from Emissions Reduction Credit Systems for Criteria Pollutants". [Background paper]

EPA. 1995. *Conservation Verification Protocols. Version 2.0.* EPA 430/B-95-012. U.S. Environmental Protection Agency. Washington, DC: The Government Printing Office.

Eto, J., S. Kito, L. Shown, and R. Sonnenblick. 1995. "Where Did the Money Go? The Cost and Performance of the Largest Commercial Sector DSM Programs." LBL-38201, UC-1322. Report prepared for the U.S. Department of Energy.

Faris, Rob, Marco Boscolo, and Theodore Panayotou. 1997. "Carbon Offsets in Costa Rica: A Case Study". Harvard Institute for International Development. Processed.

Fearnside, Philip M. 1997. "Monitoring Needs to Transform Amazonian Forest Maintenance into a Global-Warming Mitigation Option". *Mitigation and Adaptation Strategies for Global Change **2:***285-302.

Hagler-Bailly 1998. "Evaluating Greenhouse Gas Mitigation through DSM Projects: Lessons Learned from DSM Evaluation in the United States". [Background paper]

IEA (International Energy Agency) 1995. *Oil, Gas, and Coal Supply Outlook.* Paris: OECD/IEA.

IEA (International Energy Agency) 1997. *Activities Implemented Jointly: Partnerships for Climate and Development.* Paris: OECD/IEA.

Joskow, Paul L. and Donald B. Marron. "What Does a Negawatt Really Cost? Evidence from Utility Conservation Programs". *Energy Journal 13(*no. 4): 41-74.

Khanna, M. 1997. "Freer Markets and the Abatement of Carbon Emissions: the Electricity-Generating Sector in India."

Lewis, Tracy. 1997. "Inducing Real Reductions in Green House Gas Emissions from Joint Implementation Projects when Baseline Costs are Unknown". [Background paper]

Marbek Resource Consultants and RCG/Hagler, Bailly. 1994. *Guidelines for the Evaluation of Demand-Side Management Programs, Volume 1.* Prepared for Canadian Electrical Association. Boulder, CO: RCG/Hagler, Bailly, Inc.

Martin, Will. 1998. "Achieving Carbon Emissions Reductions through Joint Implementation". World Bank, processed.

Michaelowa, Axel. 1997. "AIJ- the baseline issue from an economic and political viewpoint". Http://www.northsea.nl/Jiq/complete.htm

NARUC. 1994. "Regulating DSM Program Evaluation: Policy and Administrative Issues for Public Utility Commissions." ORNL/Sub/95X-SH985C. Co-authors: J. Raab and D. Violette. Washington, D.C.: National Association of Regulatory Utility Commissioners

Palmisano, John. n.d. "Air Permit Trading Paradigms for Greenhouse Gases: Why Allowances Won't Work and Credits Will".

Parfomak, Paul W. and Lester B. Lave. 1996. "How Many Kilowatts are in a Negawatt? Verifying Ex Post Estimates of Utility Conservation Impacts at the Regional Level". *Energy Journal 17*(4):59-87.

Richards, Kenneth. 1993. *Valuation of Temporary and Future Greenhouse Gas Reductions.* Washington DC: Batelle Laboratories.

Rosebrock, Jens. 1994. *Time-Weighting Emission Reductions for Global Warming Projects.* World Bank Environment Department. Processed.

Pickrell, Don H. 1992. "A Desire Named Streetcar: Fantasy and Fact in Rail Transit Planning. *Journal of the American Planning Association, 58:* No 2: 158-176.

Saxonis, W.P. 1995. "After Ten Years of Measuring Free Riders: What Have We Learned? What Challenges Remain." *Energy Program Evaluation: Uses, Methods, and Results* CONF-950817. Chicago, IL: National Energy Program Evaluation Conference, pp. 847-858.

Swaminathan, Shiva and Samuel Fankhauser. "Measuring the Incremental Costs of Global Environmental Protection: The Case of Grid-connected Wind Turbines in Morocco" . World Bank. Processed.

SGS 1997. *Certification of "The Protected Area Project" (PAP) in Costa Rica for OCIC (the Costan Rican Office for Joint Implementation).* Oxford: SGS Forestry.

Swisher, Joel, Gilberto Jannuzzi, and Robert Redlinger. 1997. *Tools and Methods for Integrated Resource Planning: Improving Energy Efficiency and Protecting the*

Environment. Risø, Denmark: UNEP Collaborating Center on Energy and the Environment.

Nordic Council of Ministers (1996). *Joint Implementation of Commitments to Mitigate Climate Change: Analysis of 5 selected energy projects in Eastern Europe.* TemaNord 1996:573. Copenhagen: Nordic Council of Ministers.

Train, Kenneth E. 1994. "Estimation of Net Savings from Energy-Conservation Programs". *Energy, 19*(no. 4):423-441.

Vine, Edward and Jayant Sathaye. 1997. *The Monitoring, Evaluation, Reportin, and Verification of Climate Change Mitigation Projects: Discussion of Issues and Methodologies and Review of Existing Protocols and Guidelines.* Energy Analysis Program, Environmental Energy Technologies Division, Lawrence Berkeley National Laboratory. LBNL-40316.

Wang, Hua and David Wheeler. 1996. *Pricing Industrial Pollution in China: An Econometric Analysis of the Levy System.* World Bank, Policy Research Working Paper no. 1644.