

THE EVALUATION OF REGULATORY AGENCIES

Jon Stern

Centre for Competition and Regulatory Policy (CCRP)
City University London

CCRP Working Paper No 16
(March 2010)

THE EVALUATION OF REGULATORY AGENCIES

Jon Stern
Senior Visiting Research Fellow, CCRP

The paper that follows is a survey paper on the evaluation of regulation and regulatory agencies. It is a chapter (Chapter 11) in the forthcoming Oxford Handbook on Regulation edited by Robert Baldwin, Martin Cave and Martin Lodge, which is due to be published in Summer 2010 and is reproduced with the permission of Oxford University Press.

ABSTRACT

This paper presents a survey and overview of the evaluation of infrastructure regulation and regulatory agencies. It covers the development of ex ante and ex post evaluation based on cost-benefit analysis and the application of these methods to regulation in the UK, the European Union and the USA. Much of the work in this area has been on infrastructure regulation in developing countries. The paper discusses the evaluation issues involved and covers the use of case studies, econometric techniques and, briefly, randomised experiments. The paper discusses the origin, contents and lessons from the 2006 World Bank Handbook on evaluating infrastructure regulatory frameworks. The paper concludes with some general comments on the role of evaluation both for establishing the accountability of regulatory arrangements as well as for providing a process by which they and others can systematically learn from experience.

1 INTRODUCTION

Evaluation is crucial for public policy as it is how governments and agencies learn. The intellectual foundations for the economic evaluation of projects and programmes were established in the 1960s and 1970s with the rise of applied cost-benefit analysis (classic contributions include, Foster and Beesley, 1963; Layard, 1972; Little and Mirrlees, 1974). Since then, the practice of economic evaluation has developed considerably – as will be seen below.

Economic evaluation covers both *ex ante* (in US terminology, ‘before the fact’) assessments and *ex post* (‘after the fact’) and concentrates on the *outcomes* of various governmental decisions. Other types of evaluation focus on other aspects. For instance, there is a sizeable literature on evaluations of policy and management *processes*. In addition, students of government and politics have explored how and why evaluation becomes more (or less) important, and its role in the government process. This chapter focuses mostly on economic evaluations in general and *ex post* evaluations in particular. Other dimensions are important and this chapter will comment about the political and governmental role of economic evaluations, but will devote little space to process evaluations.

One point about terminology – much recent literature on economic evaluation has used the term ‘impact assessment’. This has been used to cover *both ex ante* assessments *and ex ante* evaluations. In the UK, ‘impact assessment’ is the term used by the government to denote *ex ante* appraisals of new regulations whereas in much World Bank and development country literature that term is used to denote *ex post* evaluations. This chapter sticks to UK terminology. It focuses primarily on *ex post* evaluations but will discuss *ex ante* assessments but in less detail, not least because *ex ante* assessments are often the starting point for *ex post* evaluations.

Evaluation practice was developed for public expenditure projects and programmes and has followed a relatively standard format. For *ex post* evaluations, it usually focuses on the direct and closely related outcomes of the project (e.g. the change in traffic flows from building a new road connection) and compares those with what one might have expected to occur had the project not taken place. This last is known as ‘the counterfactual’. Constructing counterfactuals is difficult and the resulting artefact is often controversial. In addition, as is discussed further below, it is virtually impossible, even if it were desirable, to construct credible counterfactuals for a regulatory agency. How could anyone possibly estimate what outputs and prices would have been for UK telecommunications products had Ofcom and Oftel never existed?

This makes evaluating regulatory agencies an especially difficult task. Even evaluating individual regulatory proposals is more difficult than evaluating specific public expenditure projects. In particular, for *ex ante* assessments, the analyses are more difficult for regulation, both technically and for political economy reasons.

In general, this chapter focuses on the economic regulation of infrastructure industries such as electricity, telecoms and water and, in particular, on the evaluation of *outcomes*, rather than regulatory *procedural processes*. Although consumer protection is the core function of both infrastructure regulators and others, such as food and drug regulators, financial regulators, and health and safety regulators, there are also major differences between these various types. Infrastructure regulation has at its heart the issue of regulating monopoly networks which have substantial economies of density, scale, and scope. This means that the *substance* of regulation is very different between the infrastructure regulators and the specialist consumer/employee protection regulators. The latter are primarily concerned with setting (and enforcing) quality

standards and dealing with market failures that arise from serious information asymmetries between consumers and companies.

However, many of the lessons from evaluating infrastructure regulators are readily applicable to the other types of regulators. This in particular applies to setting *governance* criteria and standards viz. clarity of functions, accountability, transparency, effective monitoring and effective procedures. Hence, much of the discussion of governance and a good deal of the discussion of how to evaluate regulators in Section 4 should also be relevant to all types of regulator. Many of the examples of infrastructure regulation that are cited below come from developing countries on which there has been much focus and where there are a lot of studies on which to draw, but material on the UK and other high income, OECD countries is included, where available and relevant.

The rest of this chapter first discusses evaluation methods in more detail before turning to what can be done with regard to the ex post evaluation of regulatory agencies. Section 2 discusses evaluation issues and their relevance for evaluating regulatory agencies, including the recent debate around the use of randomised trials. Section 3 discusses alternative methods that have been used in recent years to evaluate regulatory agencies including the various types of case study, econometric and other methods of statistical analysis. Section 4 considers the proposed structured case study approach in the World Bank Handbook on the evaluation of infrastructure regulatory systems that tries to combine the evaluation of regulatory governance with the evaluation of industry outcomes. This section also covers the application of that and similar methods both in the context of infrastructure regulation and of merger remedies in competition policy. Section 6 provides some short concluding observations.

2 EVALUATION METHODS AND THEIR APPLICATION TO REGULATORY ENTITIES

Standard methods of economic evaluation apply cost-benefit analysis to proposed or realised government interventions.¹ Most such evaluation – ex ante or ex post – focuses on *partial equilibrium* analysis. The main objective of this is to estimate the impact of those directly or closely affected by the intervention. The analysis is ‘partial’ in that it excludes wider impacts on relative prices, feedback effects to and from other economic sectors, or the final long-run net impact on the economy as a whole.

For instance, a partial equilibrium ex post evaluation of a road widening project will estimate the changes in traffic flows, time savings, environmental pollution, etc. arising on the widened road and the neighbouring area. It will then compare (a) the benefits (positive, one hopes) and (b) the costs of the road widening with: (i) what was expected in the proposal for building the road (including forecasts of traffic flows and consequentials both with and without the widening of the road); and (ii) traffic flows and environmental pollution levels, etc. in the years before the road widening took place.

The key point is that the project can be evaluated in this partial equilibrium way because it is a *marginal* project. It is a marginal project in that it only adds a very small amount of extra national road capacity; and, hence, does not significantly change the relative prices of time, petrol and diesel, lorry freight rates, road construction costs, etc.

Partial equilibrium cost-benefit based evaluations run into difficulties the larger and more widespread is the government intervention in question. For instance, WTO-induced changes in

tariff rates for a wide range of internationally traded goods and services are bound to induce significant and wide-ranging change in relative prices – and, indeed, they are intended to do so. Hence, evaluation, ex ante or ex post, of trade reform packages requires *general equilibrium* analysis. These are ‘general’ in that they specifically include all expected impacts on relative prices, feedback effects to and from other economic sectors, or the final long-run net impact on the economy as a whole.

General equilibrium based evaluations are much more technically complex than partial equilibrium approaches. They require relatively sophisticated modelling techniques, and the results depend heavily on the economic and mathematical assumptions underlying the model used. Perhaps most importantly, it is also very difficult to analyse the *mechanisms* by which the impacts are expected to arise or have arisen. This last point makes general equilibrium based methods particularly unattractive for the evaluation of regulatory agencies and/or regulatory proposals and I will not discuss them further in this paper.

2.1 The Role of Economic Evaluations for Government Decision Making

Economic evaluations are virtually always promoted and managed by the most central entities in government. To use the UK as a case to explore this point, the practice and use of evaluation has been very much the responsibility of the Treasury and the ‘Green Book’ which sets out the rationale and methodology of evaluation practice has always been a Treasury publication. This was first issued in 1973 and has gone through various updates with the latest version (the 7th) published in 2003.² Other countries produce similar guidance documents and they are always written, and issued by the ministry of finance or the budget, or similar central public expenditure and/or budgetary co-ordinating agency.

The driving force for the Green Book since its third (1982) edition has been to establish a common methodology within which to consider public expenditure bids by government departments and to provide a framework within which claimed special factors can be contained. The same applies to similar documents issued in other countries. For departments, it provides a common framework within which each can weigh up and submit expenditure proposals, for e.g. road building relative to railway subsidies or new hospitals relative to public health advisory expenditures. For the Treasury, it enables a much more straightforward and contained way in which to consider public expenditure proposals within and, more importantly, between departments.

All decisions on public expenditure have, of course, a large and important political component – and the political elements will vary in content and importance over time, and between governments. Nevertheless, the ex ante assessment process provides a major way in which the political elements can be identified and controlled for by the central budgetary authorities even if they remain present.

Not surprisingly, the role of project assessment and evaluation tends to develop – and increase – at times when public expenditure growth needs to be reined in (at least in the perception of the key budgetary decision makers). This pattern has also been apparent with regulatory evaluation (including recent discussions of ‘regulatory budgets’) where the rapid and,

to many, controversial growth of regulatory interventions in the UK and the EU (European Union) over the last decade has been a major driving force for the growth of regulatory evaluations.³ This has been greatly fuelled by the claimed ‘burdens’, particularly ‘burdens on business’, from much consumer and citizen protection regulation.

Within the EU, there has been strong pressure for more evaluation of EU programmes and more specifically of EU regulatory initiatives – primarily ex ante assessments but also ex post evaluations. This has, in large part, been led by countries like the UK and other (mainly northern) EU member states who want more focus on outcome assessment and evaluation, and who want to see more economic and commercial justification of EU-wide regulatory proposals, particularly EU harmonising proposals. For other member states, the focus is more typically on the assessment of the legal quality of regulation and its harmonisation. Similar issues and pressures arise in explicitly federal countries like the US, Germany, and Australia.

The political and budgetary pressures are initially always for ex ante assessments. However, once these have been in place for a while, the pressures for ex post evaluations grow. That is partly to see how well the initial assessments performed as guidance and also to learn lessons for the future. One obvious political tension is that the Treasury and similar central entities have even more incentive to try and codify. Conversely, departments and sub-agencies have the incentive to try and establish special treatment for things of particular importance to them within the general framework but without explicitly challenging the reputation or role of the framework as a way of organising an ordered discussion for public expenditure (and regulatory) decisions.

Ex post evaluations and revisions to central guidance provide one forum within which these debates can take place – typically between experts from and consultants to the various interests. This too is apparent in the debates between the EU Commission and Member State Governments. It also arises within international organisations such the OECD and the World Bank.

The World Bank has always taken ex post evaluation very seriously and has an Independent Evaluation Group that conducts regular evaluations of programmes and strategies. This is not least because lenders to the World Bank and the regional multilateral development banks require a high level of accountability for their aid, like all bilateral and multilateral aid donors. A major issue for the World Bank is always how far approaches should or should not vary between international regions and countries. One example of this was the Bank’s 1993 evaluation of the East Asian development model which resulted in a 402 page volume evaluating the differences and outcomes between the ‘East Asian’ and the World Bank development policies and whether or not lessons from the East Asian approach should replace some of the ‘Washington Consensus’ based policy recommendations. This is one of the most high profile ex post evaluation exercises undertaken by the Bank. Since then, there have been hundreds which have covered issues ranging from further high-level evaluations of growth and development strategies to micro evaluations of detailed interventions, such as a 2009 study of programmes to improve children’s reading skills in Peru.

Major pressures for evaluations often arise when there is a perceived need for a strategic review of a line of policy e.g. where a particular approach is seen not to have delivered what was expected or hoped. This applies both at national level and more widely. The boundary between this and an investigation of what went wrong in a disaster – and who should be blamed – can be a fine one but there is a distinction and it is an important one.

In a later section, I discuss the World Bank *Handbook for Evaluating Infrastructure Regulatory Systems*. This was commissioned because expectations of independent regulatory agencies for electricity and other infrastructure industries in developing countries had not delivered what was hoped or expected. In consequence, the question arose as to whether this was because the policy model was sound but had been badly applied or was inherently flawed.

However, this general issue applies more widely. For instance, the major 2003 World Bank ex post evaluation of its water and sanitation programmes was undoubtedly triggered by a feeling that the World Bank water reform programmes had been disappointing and that a fundamental review was needed. That review led to a major (and continuing) rethink of World Bank policy for water and sewerage, and the role of private investment in developing countries.

2.2 Standard Cost-Benefit Based Evaluation Methods

There are many government issued and other manuals on the evaluation of public expenditure projects. The standard UK example is the Treasury 'Green Book,' as already noted.⁴ The latest (2003) version of the Treasury Green Book includes guidance on the use of ex ante appraisal methods for proposed new regulations, which are discussed under the description of RIAs (Regulatory Impact Assessments). In 2008, RIAs were relabelled and issued with updated guidance as IAs (Impact Assessments).⁵

Manuals like the Green Book are intended to be used by all government entities in the process of managing public expenditure and regulation programmes, and projects. The core of the approach recommended in the Green Book (p.3) is the ROAMEF cycle that is constituted by the following stages: rationale, objectives, appraisal, monitoring (during the implementation stage), evaluation and feedback (that then feeds into a reconsideration of the rationale).

Under this approach (which is also standard for other OECD countries), any proposal for a new expenditure or regulation project should start by setting out a general Rationale. For a potential regulation to stop selling cigarettes to people under 16, the rationale would be something like the objective of reducing mortality and morbidity in the population from smoking induced illnesses. The Objectives would be framed in terms of some specific numerical targets (e.g. to reduce teenage smoking by some specified percentage). The Appraisal would consider all relevant options both regulatory and non-regulatory (e.g. regulations imposed on retailers, identification materials for under-21 year olds wanting to buy cigarettes, higher cigarette prices, advertising restrictions, etc.). The costs and benefits of these should then be compared with each other and with a 'do-nothing', status quo option. If one of the options is chosen for implementation, the resulting process and effects are Monitored and, at some appropriate point, an ex post Evaluation may be carried out on the regulations or on some particular aspects of them. The results of the Evaluation are then used as Feedback to help develop regulatory improvements in this and or other regulatory areas.

Several points are worth making on this model as it applies to regulation and its evaluation in the UK. First, the methodology is intended to apply to new regulatory agencies as well as new regulations. The Appraisal part of that is extraordinarily difficult and the Evaluation part virtually impossible because: (a) they are emphatically not *marginal* changes; and (b) because so much else is happening to affect outcomes beside the creation of a new regulatory agency. For instance, creating a new infrastructure regulator is typically part of a large-scale programme that includes unbundling a highly vertically integrated company and commercialising it e.g. either by privatising it or introducing private capital on a large scale.

Second, ex ante appraisals of new regulations have been made mandatory in the UK (and now in the EU) both for government departments and for many regulatory agencies. The latter includes all the UK utility regulators and the FSA (Financial Services Authority). The UK National Audit Office (NAO) has published several reports on RIAs. These have consistently pointed out the poor quality of RIAs carried out by government departments and the latest,

covering RIAs in 2005-06, also points out how little ex post evaluation has been done by UK government departments on regulations that have been introduced after an ex ante appraisal.⁶

Third, RIAs as carried out by independent infrastructure regulatory agencies (the ‘economic regulators’) have a markedly better but still uneven record. This was reflected in a 2007 NAO report.⁷ This showed examples of very good practice (e.g. the Ofcom appraisal of restrictions on television advertising for food and drink products to children in 2006⁸) but also several examples of weak practice. The NAO 2007 report was also critical of how much less ex post evaluation had been carried out by the utility regulators than they would have expected, as well as how little was being planned.

Fourth, neither the Green Book nor official ‘better regulation’ guidance explicitly discuss ex post evaluations of regulations. The NAO (National Audit Office) has carried out investigations of specific regulatory interventions by various regulators, including the ones discussed above.⁹ Many of these are interesting and produce useful and important results but they do not use formal evaluation methods.

The conventional formal approach to ex post regulation as used in the UK, the US, EU and elsewhere is to compare cost and benefit outcomes with a counterfactual as well as with what would have been expected on past trends. The 2003 Green Book defines a counterfactual as, ‘what would have happened if the activity under consideration had not been implemented’ (HM Treasury Green Book, 2006: 47). However, it also suggests that ex post evaluations: use more than one counterfactual (i.e. ‘alternative outturns given different states of the world or different management decisions’);¹⁰ and include the use of control groups (e.g. for vocational training programmes). This perspective is geared to public expenditure programmes and it is much more difficult to apply to regulatory interventions. That is most obvious on the use of control groups in ex post evaluations which will be discussed in more detail below.

The discussion above has focused primarily on UK practice in regulatory appraisal and evaluation but the EU has adopted a similar approach. However, their approach is less firmly based on economics and cost benefit analysis, and as yet, is less well-developed so that it has been even more heavily criticised than the UK methods.

2.2.1 *Political Economy Limitations with the Standard Evaluation Model*

All countries and institutions seem to find it much harder to appraise and evaluate regulations than evaluating expenditure projects or programmes. They also find evaluating regulatory agencies particularly difficult. Some of this difficulty seems to arise from the technical difficulties of applying cost benefit analysis to regulations and, even more, to regulatory entities, but there are also major political economy and pure political issues.

As regards political economy, the cost benefit approach assumes a benevolent policy maker or planner whose objective is to maximise social welfare – i.e. a disinterested public service technocrat. This assumption may be reasonably appropriate for staff working in a regulatory agency which has readily enforceable legal obligations both to behave fairly and to demonstrate publicly the reasons for its decisions,¹¹ but it is much less likely to be appropriate for government departments. For the latter, it is to be hoped that public service motivated specialists will give objective advice internally before decisions are made but that advice will inevitably be affected, to some degree, by the prevailing political climate and the ambitions of the staff.

For government decision makers (most obviously Ministers), the published ex ante assessments of regulations like RIAs are documents *justifying* the decisions that have been made. Similarly, ex post evaluations that expose failures are usually highly unwelcome. Hence, there are major disincentives on departments/ministries to publish objective assessments and

evaluations, particularly ones by *their staff*. This seems to be a major reason why policy observers and academics, journalists and others are so uniformly critical: (a) of the quality and apparent lack of impact of the ex ante regulatory appraisals emerging from government departments/ministries (and the EU); (b) regularly point to the absence of thorough-going ex post evaluations by them (Hahn, 2008); and c) of the limited relationship between the quality of evaluation exercises and the quality of regulation per se.

The bleak picture above is, however, partially offset by the activities of policy audit agencies like the NAO and GAO which, typically, are agencies of – and report to – the legislature rather than the executive. In addition, the need of government departments/ministries to learn, means that they frequently commission external academic or other consultants to evaluate their programmes.¹² The resulting reports may or may not be published - and, if they are published, it is (not surprisingly) usually with a press management/public relations gloss giving the department's desired interpretation of the results. However, as yet, at least in the UK, this use of external ex post evaluation is much more developed for public expenditure projects and programmes rather than for regulatory interventions.

For genuinely independent regulators (i.e. regulatory agencies not funded from general budget revenues and with security of tenure of regulatory decision making staff), the same disincentives exist but they are offset by legal obligations on disclosure and justification. Regulatory agencies can be taken to court and their decisions formally appealed against if they act in ways that cannot be justified. The same does not apply to government ministries/departments (at least in Europe). Being taken to court in this way can be very embarrassing and lost cases can rapidly destroy the reputation – and hence the credibility and viability - of the regulatory agency.

In consequence, the incentives for making and publishing technically sound, and objective appraisals is much higher for genuinely independent regulators lodged in a well-functioning legal framework. Of course, this does not apply to circumstances or countries where regulatory entities are not genuinely independent in the way described above. They may not be independent *either* because the legal basis on which they operate does not give them de jure independence *or* the legal and political framework in which they operate means that, in practice, they are not de facto independent (see Levy and Spiller, 1996; Stern, 1997; Stern & Holder, 1999; Brown, Stern, and Tenenbaum, 2006). This is one of the main reasons why evaluation of regulatory *governance* is a fundamental aspect of the regulatory evaluation of outcomes, as I will discuss more fully below.

2.3 Evaluation Approaches: Counterfactuals, Control Groups, and Comparators

This section considers the relative pros and cons of classic ex post evaluation with counterfactuals, relative to the use of evaluations using control groups and experimentally designed evaluations.

2.3.1 Ex Post Evaluation with Counterfactuals

Constructing credible counterfactuals is difficult and controversial. For example, one of the best known ex post evaluations of infrastructure industry reform and regulation is the Newbery and Pollitt social cost benefit analysis of the privatisation of the England & Wales CEGB (Central Electricity Generating Board). This study was the subject of a later re-examination by Horton and Littlechild (Newbery and Pollitt, 1997; Horton and Littlechild, 2002). Newbery and Pollitt had

estimated the net benefits at around £4-10 billion (maybe £13 billion) but with net costs to consumers from the first few years of privatisation. Conversely, Horton and Littlechild (provisionally) concluded that the net benefits were around twice as large as the Newbery-Pollitt estimates (around £20-25 billion) and with substantial net benefits to consumers right from the immediate post-privatisation period.

The differences in the estimated benefits and costs are large but not atypical for such analyses. The Horton-Littlechild re-analysis shows clearly how sensitive the results of counterfactual based analyses can be to changes in the choice of assumptions. They demonstrate clearly that such evaluations are at least as much of an art than a science – a scientifically informed art; but an art, nevertheless. In this case, as in many others, the readers are left to conclude which estimates they incline to on the basis of judgements concerning the alternative sets of assumptions. However, evaluations of economic regulation are not always so difficult.

2.3.2 Control Groups and Experimental Design of Evaluations

To reduce the degree of judgement required, the use of control groups has frequently been recommended to make evaluation more of a science than an art. The ‘gold standard’ in this area is the use of trials with random assignment of participants between treated and control groups. In socio-economic policy, there was a strong move in this direction in the 1980s in the labour economics and social security areas, including various US negative income tax and training programme experiments. Underlying this push was the inspiration drawn from pharmaceutical testing where new medicines are tested with one group of people being given an active medicine and the control group being given a placebo. For such medical trials, it is usual to involve ‘double-blind’ procedures under which neither the administering doctor/researcher knows whether the patient has been given the active medicine or the placebo. That is not possible for economic intervention, e.g. for a training programme for the long-term unemployed. That in itself raises problems and has generated major debates for various evaluations as to whether or not the random assignment was genuinely random. For instance, there have been suggestions that, at least in some programmes, those administering the programmes have, on occasion, shifted a control group person to the programme group because they decided the person was particularly deserving or likely to benefit. Similarly, participating institutions involved in providing the training or similar programmes volunteered and were not randomly chosen (Heckman, 1992).

In development economics, there has been a huge surge of interest in recent years on the use of random assignment evaluation models for development programmes and strong claims that they result in ex post evaluations that are inherently superior to those based on standard counterfactual analysis. The leaders of this school of thought are Anghit Banerjee and Esther Duflo (see Banerjee and Duflo, 2008). Although their papers are high quality technical pieces of work, the results are highly specific and just do not generalise. Experiments are experiments and the results of them are subject to positive biases, not least because all concerned (researchers, demonstrators, and participants) know that they are in a demonstration project. Indeed, the results of the modelling typically raise as many questions as answers in terms of interpretation and applicability – and largely the same questions as in the previous generation’s work on experimental approaches to labour and training policy. Not least, the random assignment evaluation model is not well-g geared to explaining how and why results occur (Deaton, 2009).

The random control model evaluation approach may have some merits for evaluating programmes to improve developing country programmes eg to raise low income country schooling rates or anti-malarial bed-net programmes, although even there the implications for policy are primarily very local. However, in its pure form it is totally useless for evaluating new

regulations, let alone new regulatory institutions. Hence, the purist version of the control group revival would just rule out conventional ex post cost benefit analysis of regulation or regulatory institutions. This is plain silly and, fortunately, is not what happens in real world policy debates.

In contrast with the random control approach, Ravallion (2008) discusses how ‘evaluative research’ has been successfully used in practice for development policy design. He discusses some of the problems with applying the results of experiments, e.g. scaling-up issues. He also discusses at some length the productive experiences from evaluations based on comparisons of scheme outcomes in different Provinces for the design of Chinese economic reform (including, regulatory reform). He describes the Chinese approach as ‘seeking truth from facts ... [with] a high weight put on demonstrable success in actual policy experiments on the ground’ (Ravallion, 2008: 2). Rodrik describes this as, ‘seeing whether something worked’ (2008: 26). It compares results from local (non-randomised) experiments taking as counterfactuals: (a) changes in outcomes in experimental areas pre and post the experiment; and (b) comparison with outcomes in areas without experiments.

Like development economics, the certainties of 15 years ago regarding infrastructure industry reform and regulation have come under increasing question – particularly for developing countries and countries with weak institutional frameworks. We know that regulatory agencies and methods do not readily transfer from the US or Europe to Africa, or Asia. Commercialisation still seems superior to the alternatives, but quite how that is achieved and what regulatory support is necessary and by what type of institution is much debated. There are clearly no blueprints.

As a result, Rodrik argues for an ‘open-minded, open-ended, pragmatic, experimental approach’ to evaluation and policy making in development economics. This contrasts with the prescriptive approach ascribed to the Washington Consensus of the 1980s and 1990s – ascribed fairly for its institutional recommendations, if somewhat unfairly on other recommendations. Rodrik has long been a critic of the Washington Consensus arguing that while there may be considerable consensus on economic *ends*, the institutional and policy decisions *means* vary considerably between countries. Hence, the emphasis must be on locally based diagnostics, including evaluation. Rodrik (2008) cites approvingly the Ravallion paper on evaluation and the Chinese experimental approach based on learning from local/Provincial experiments, and the practical lessons from their practical (non-randomised) evaluation approach.

This approach to evaluation is echoed in comparisons of infrastructure and other regulation across US States. There have been many case studies and some formal econometric studies (e.g. Besley (2002) that consider US infrastructure state regulators and differences in decision-making depending on whether regulators are elected rather than appointed). This provides a form of evaluation which is clearly in the same comparative perspective as the control group approach but much less formally constrained and, hence, much more useful for general lessons. This type of approach is also spreading within the European Union with increasing benchmarking and evaluation of regulatory and other institutional operations by the EU Commission and academic researchers.

Such a more open-minded approach has also spread to infrastructure regulation per se, particularly for developing countries and especially for developing countries with weaker legal and regulatory institutions. Hence, the longest chapter in the 2006 World Bank Handbook for Evaluating Infrastructure Regulatory Frameworks (Brown, Stern, and Tenenbaum, 2006) was on intermediate and transitional regulatory systems i.e. what one might work best in countries like Ghana or Senegal, Sri Lanka, or Thailand. The emphasis there was on identifying ‘good fits’ rather than trying to impose best OECD practice. ‘Good fits’ were identified as regulatory arrangements that: (i) met immediate needs; (ii) were operational given the concerns and

limitations within the country and sector; (iii) and were capable and incentives for progressive enhancement and improvement.

Under this heading, Brown, Stern, and Tenenbaum advocated specific additional criteria for evaluation of these intermediate and transitional regulatory frameworks besides those appropriate e.g. in OECD countries. These demand that the current situation is established and that consideration be given to:

- 1) Whether and how far the observed regulatory system has moved to improve its regulatory practices relative to previous arrangements (using meta-principles, plus principles, and standards);
- 2) The contribution of the regulatory arrangements and decisions to the quality of targeted industry outcomes achieved;
- 3) How the regulatory arrangements and industry outcomes compare with those in similar countries;
- 4) The internal incentives and pressures that are likely to improve regulatory processes and industry outcomes, or likely to impede or retard progress, or threaten the viability of current regulatory arrangements, or both (see Brown, Stern and Tenenbaum 2006: 92).

This approach to evaluation is of the same family as those advocated by Rodrik and Ravallion and again arises to a considerable extent from the hard lessons learned over the last 15-20 years about what regulatory frameworks are and are not effective, in what contexts and why. How these criteria can be employed in actual evaluations will be demonstrated in Section 4 below.

3 CASE STUDY AND ECONOMETRIC APPROACHES TO REGULATORY EVALUATION

This section considers two approaches towards regulatory evaluation, namely case studies and econometric-based approaches. Excluded from this discussion are attempts at benchmarking of regulators as another method of appraisal frequently used to compare regulators in general and infrastructure regulators in particular. However, it typically involves comparisons of regulatory governance and/or resource inputs rather than comparisons of performance on outputs or outcomes. For that reason, they are not discussed in this context.

3.1 Case Study Evaluations of Regulators

Case studies primarily provide potential hypotheses, the generality of which needs to be tested in some other way. This is certainly the case for one-off evaluation case studies. Hence, evaluations by case study are complementary to more formal econometric and similar evaluations but are not a substitute for them. Case studies are probably most helpful to provide a context within which policy can be discussed and least helpful in establishing causation, particularly causation across countries.

Space prohibits a comprehensive survey of the numerous case studies that encompass regulatory agencies and individual regulations. For outcome evaluation purposes, the most useful case studies are those which are set up to provide directly comparative studies of similar regulatory initiatives in different places. In particular, there are major benefits of using a common regulatory framework and a common questionnaire (or interview and data collection framework). A good example of a set of comparative case studies of the type recommended is the study of five Latin American and African capital city water supply and sanitation reforms co-ordinated and edited by Mary Shirley (Shirley, 2002). Also for water and sewerage, Ehrhardt et. al. (2007) draw on a similar set of case studies in five other places. These are evaluations of industry and institutional performance written with the explicit goal of providing evidence as to what is useful, where, and why – and what is damaging, where, and why. Hence they take an ex post evaluation perspective even if they are not formal evaluations with explicit counterfactuals. As such, they are clearly moving in the direction of the approach recommended by Brown et al (2006) and away from the classic one-off detailed case study.¹³ In contrast, the seminal Levy and Spiller (1996) book of comparisons of telecom regulation in Argentina, Chile, Jamaica, the Philippines and the UK is well-known as a set of comparative studies but they are more a linked set of economic histories than a set of evaluations. For case studies to provide evaluation or regulatory agencies, and regulation that can be usefully generalised, it is necessary *either* to adopt and take forward the comparative approach as used in the Mary Shirley book *and/or* to bring in more aspects of the classic counter-factual evaluation.

3.2 Econometric Evaluations of Regulatory Impacts

There is now over ten years of experience with using econometric models to establish the impact of infrastructure regulation on, first, telecoms and second, electricity industry outcomes. Various statistical techniques are utilised (primarily econometric techniques based on variants of multiple regression analysis) to examine whether various formal and informal characteristics of the regulatory system have produced positive or negative effects on the economic performance of the sector.

The data for these studies usually come from published information and questionnaires sent to the regulators in different countries. Typically, the studies try to determine whether certain regulatory characteristics or combinations of characteristics (such as institutional independence, existence of a regulatory statute, or type of tariff-setting system) have had positive or negative effects on different dimensions of sector performance (such as, levels of investment and capacity utilization). The studies attempt to use real world data to test general propositions on the potential economic effects of regulation. However, they are neither designed to evaluate specific regulatory decisions nor to provide detailed recommendations on specific reforms. This does not, however, imply that the studies are irrelevant to the real world of regulation. Presumably, policy makers will benefit from knowing whether different dimensions of regulatory governance (for example, an independent regulator) and substance (for example, cost of service versus price cap regulation) are associated with increases in infrastructure industry investment, productivity, and performance.

Quantitative answers to these fundamental questions can only be found from econometric studies of this type, whatever qualifications may be attached to specific studies. With the growing

availability of ‘panel data’ (comparable data for a good number of countries for a number of years), the quality of the econometric studies has greatly improved and their results have become more comparable and more consistent. In general, the more recent studies increasingly confirm the view that good regulation (as defined by the good governance characteristics) improves investment and productivity performance in developing countries both in telecoms and in electricity.

These cross-country statistical studies are not designed to provide an in-depth review of the performance of a single country’s regulatory system (or some specific elements of the system). In addition, they provide corroborative evidence where causation may be strongly indicated but cannot be conclusively demonstrated. Hence, cross-country econometric studies and single-country evaluations complement each other, but they are quite different in both objectives and methods, and the ‘evaluation’ provided by the econometric studies is much more general.

This literature focuses primarily on the impact of the existence and governance quality of regulation – along with the changes in market liberalisation and privatisation – on investment and efficiency levels in developing country telecom and electricity industries. The main reasons for the focus on developing countries are: firstly, infrastructure development is a major issue in development policy; and secondly, there is the issue that improvements in regulation are intended to increase the *supply* of capacity. For developing countries, there is no question that supply is seriously inadequate to meet the demand for electricity, telecoms, water and sewerage whereas, in developed countries privatisation and regulation were sometimes introduced to help reduce incentives for excess capacity and investment (viz. UK electricity). It is extremely difficult to use econometric studies to establish supply impacts if the data sample used combines a number of countries with excess capacity supply together with a number of countries with inadequate capacity to meet demand.¹⁴ In consequence, the following sticks primarily to approaches to evaluating the impact of regulation on infrastructure investment and efficiency in developing countries.

The now standard best practice econometric procedure is to take a panel data set (e.g. around 20 countries with 10-20 years of data for each) and to estimate an econometric model for physical investment (and separately for efficiency) in telecoms or electricity that allows for both observable and unobservable country-specific effects.¹⁵ The models also include terms for regulatory quality, industry liberalization, and competition, as well as country governance characteristics, and other relevant control variables (e.g. income growth, urbanization rates, relative prices, etc). Variants of this model have been estimated for telecoms (fixed and mobile), and for electricity (generation and distribution).

The power of this method arises from the fact that countries in the sample established infrastructure regulatory agencies at different dates. The same applies to liberalization and privatization (where relevant). Also, there are a few countries which have not established a telecom or electricity/energy regulator by the end of the observation period. Combining these factors provides considerable power to the panel data economic techniques which means that comparisons can effectively be made across countries over time periods, as well as, within countries over lengthy periods.

Initially, research focused primarily on the impact of ‘independent regulatory agencies’ on investment and efficiency outcomes. This has been increasingly replaced by a broader interest in the impact of regulatory governance, as measured by a number of indicators. However, as the number of indicators increased, the less clear became the conclusion as to which factors were the most crucial for achieving desirable outcomes. Subsequently, there has been a move towards grouping indicators (see Guiterrez, 2003 who shows a significant relationship between regulatory governance and investment, and efficiency growth in Latin America and the Caribbean).

Montoya and Trillas (2009) provide evidence on de facto practical aspects of the quality of telecom regulatory governance (viz. the security of tenure of regulatory commissioners). In their model, better performance on de facto governance significantly improves the association between regulatory quality and fixed line telecom penetration rates relative to the model that just relies on the quality of governance in the relevant law. The results on the tenure of regulatory commissioners (or similar), is very consistent both with case study work on regulatory governance and with research on the effectiveness of independent central banks. It is particularly important as to whether regulatory commissioners remain in post following a change of government. Similar results have been established for electricity. Cubbin and Stern (2006) showed positive effects of regulation – in particular once an agency had been established for three-five years – on investment in generation in the case of 28 developing countries (also Andres et. al., 2008). In the developed world too, there has been major growth in work using econometric methods to appraise the effectiveness of specific pieces of telecom or electricity regulation. This is particularly the case in the US where there exists comparative data by individual state and Federal regulators and this is growing in the EU. Many of these studies are interesting but, frequently, they are investigations (varying from the relatively academic to the more quasi-advocacy) of proposals for some or other regulatory reform. Hence, they are at some remove from ex post evaluation and are not discussed further here.

Econometric analyses face a number of problems, however. Most of all, regulatory agencies neither operate in a vacuum nor are they introduced randomly. Governments with broad ‘good’ quality of governance are more likely to establish credible regulatory governance, although some institutions are clearly established to merely please the loan conditions of international organisations. This problem is what econometricians call an ‘endogeneity’ problem. In other words, the modelling has to take seriously the issue that the existence of an infrastructure regulatory agency – and its governance quality – are systematically related to *country* governance and other factors in the model to be estimated. This issue has been much explored e.g. by Gutierrez (2003) as well as Maiorano and Stern for telecommunications and by Cubbin and Stern for electricity generation.¹⁶

4 THE WORLD BANK HANDBOOK ON EVALUATING INFRASTRUCTURE REGULATION

In 2006, the World Bank published the *Handbook for Evaluating Infrastructure Regulatory Systems*. The Handbook (of which I was a co-author) was originally planned as a relatively short monograph on the evaluation of – and set of guidelines for – regulatory governance and regulatory outcomes for electricity. It emerged as a full-blown book not least because, for very good reason, it decided to pursue in detail the issues around the design and evaluation of ‘transitional and intermediate’ infrastructure regulators i.e. infrastructure regulatory arrangements for countries unable (because of human resource or other constraints) and/or unwilling to establish ‘independent’ regulatory agencies of the type familiar in the UK, US, EU and other richer OECD countries. The chapter on this subject was much the longest in the Handbook.

Discussing this Handbook in some detail addresses many issues that are at the heart of this chapter. It represents current best practice in non-econometric, case study-type methods of evaluating infrastructure industry regulatory agencies and decisions. Indeed, a very similar approach has been suggested for evaluating the performance of competition agencies regarding

merger decisions and other types of regulation. Furthermore the Handbook also provides insights into dominant assumptions that characterise current thinking, namely (i) that a commercialised environment in which private investment dominates is preferable, and (ii) that traditional vertically integrated monopolies are vulnerable to corruption and (iii) that institutions and their design are important in facilitating investment and therefore economic development.

The reason why the World Bank should be at the forefront of developing evaluation methods is that they are accountable to such a wide range of bodies. The list includes both donor/lending countries and borrowing countries, as well as a wide range of development professionals, including both the Bank's own staff and NGOs, and their staff. Hence, there is rather more pressure on the World Bank and similar agencies to demonstrate that their policies and grant/lending decisions are well-founded. Indeed, there is almost certainly rather more pressure on multilateral aid and development agencies than there is on national governments whose existence and funding does not depend on the goodwill and actions of other countries and their governments.

4.1 The Context and Purpose of the Handbook and Regulatory Evaluations

By 2004, the World Bank and its infrastructure specialists were well aware of the plethora of material on regulatory governance principles and on regulatory benchmarking. But, it was also clear that there was very little information on whether and how far the rapid growth in the number of infrastructure regulators over the previous decade had or had not significantly improved infrastructure industry outcomes, let alone *how*, and how it could be improved. Member states were interested in devising a system that would allow for regular 'check ups'. Furthermore, this Handbook also emerged in a context of disappointment. Most of all, while in telecommunications the experience with regulatory institutions had been positive, no such clear pattern had been discernable in the case of electricity. Indeed, the agency model had proven difficult to embed in particular political systems and had witnessed considerable resistance. This represented a near pendulum-swing from the early 1990s regarding the effects of establishing supposedly independent regulatory agencies and a lack of consideration as to the limits of regulatory activities in transition and developing countries. For example, this author remembers distinctly how World Bank staff discounted the political problems of raising household electricity prices by a third once they had been imposed by a regulator. In short, the call for a Handbook reflected a growing realism as to the possibilities of regulation.

The Handbook gives guidance on how the possible elements of transitional regulatory systems should be evaluated. The important point is that they need to be evaluated in terms of *both* how well they deal with current circumstances in the relevant country and sector *and* whether they provide a route and incentives for moving toward significant and sustainable improvements in regulatory practices and sector outcomes. The ultimate goal is a best-practice regulatory system—a regulatory system that transparently provides investors with credible commitments and consumers with genuine protections' (Brown et. al., 2006: 9).

The 'best practice' benchmark for evaluating regulatory agencies was defined as the 'independent regulator' model and the Handbook had a core chapter plus a long Appendix setting out in detail best practice regulatory governance standards as they would apply to that model. However, the Handbook was also careful to define the benchmark independent regulator model as including a wide range of variants on this, including mixed contract and regulator approaches and others (Brown et. al., 2006: 51-4). In addition, best-practice regulatory options were clearly recognized for many developing countries as at best a very long-term objective.

Evaluation was designated a central role in helping to develop and improve infrastructure regulatory practice in the medium term in countries where the independent regulatory model was infeasible or inappropriate for many years to come. This would apply in particular to countries like China and Vietnam where there is no tradition of the separation of powers but where there was a considerable degree of willingness to operate infrastructure industries on a commercialized basis within a generally market-oriented framework.

The Handbook provides an analysis and evaluation framework for countries where regulatory and other institutions were weak (or where the ‘independent’ regulator model, was unacceptable for political and/or legal reasons). However, it is much less clear that it is useful or relevant in countries where commercialization of infrastructure industries is unacceptable (e.g. Venezuela under the Chavez government and others). As discussed in the introduction to this section, the Handbook model assumes the commercialization of infrastructure industries (and the enforcement of contracts and law) as necessary prerequisites for the effective development of any effective regulatory entity.

Hence, the Handbook approach is arguably of little use *at least as a learning tool* in countries and industries where commercialization is doubtfully accepted or highly constrained like the Indian electricity industry or for most developing country water industries. It may still be relevant as a *diagnostic tool* but such diagnoses will inevitably feature on the problems with and impediments to commercialization, whether from government policy, regulatory or from deep-seated political ideological forces.

4.2 The Handbook Approach to Evaluating Infrastructure Regulatory Entities

The most important developments that the World Bank *Handbook* has made to the evaluation of regulation and regulatory agencies are:

- (i) It links the evaluation of regulatory governance and the outcomes of regulated industries;
- (ii) It provides workable evaluation strategies within the partial equilibrium economic evaluation framework that allow for the impact of other factors as well as regulatory interventions and, also, suggests a method of analysis for decisions on a *marginal* basis, following standard cost-benefit principles, as outlined in Section 2.
- (iii) Suggests ways of evaluating intermediate and transitional regulatory agencies in terms of their previous and potential development, as well as their outcomes to date.

Note that the above is written in terms of regulatory agencies per se. The *Handbook*, of course, deals with just the economic regulation of infrastructure industries; indeed, it focuses almost exclusively on electricity. However, the evaluation methodology can be readily modified to cover other infrastructure industries (and has already been so adapted, as will be shown in Section 5). Further, there seems no reason why the methodology cannot be applied to the ex post evaluation of any other type of regulator.

The key to the *Handbook* methodology is the belief that higher quality regulatory governance is strongly expected to improve the quality of regulatory decision-making and hence, other things equal, the outcomes of regulated industries. To quote the Handbook again, ‘... the recommended approach is to look at specific elements of the regulatory system that relate both to regulatory

governance and substance, and to assess whether they help or hinder sector performance (Brown et. al., 2006: 43).

It is not immediately obvious how this relates to the classic ex post evaluation with counterfactual. However, it clearly does so even if somewhat indirectly. Firstly, the *Handbook* clearly (and strongly) recommends the collection of quantitative data where relevant, particularly for in-depth evaluations; and, secondly, a footnote on counterfactuals states ‘Given the more limited objectives of identifying, in qualitative terms, positive and negative contributions, construction of an explicit general counterfactual is not proposed. *However, the evaluators, in developing any recommendations for regulatory reform, will implicitly be assessing what might otherwise have happened if the regulator had taken a different view or made a different decision.* [My emphasis] This is a much less formal and more limited notion of a counterfactual than the one used in the academic studies to assess overall reform programs’ (Brown et. al., 2006: 53, footnote 17).

As will be shown in Section 5 below, applying this methodology in practice typically involves:

- (a) Comparing pre-and post key decision outcomes; and
- (b) Where possible, comparing outcomes with those in a relevant comparator country where a different regulatory choice was made.

However, both pre- and post-outcome comparisons, as well as, comparator situation comparisons are often used to construct counterfactuals e.g. in the road expenditure project evaluations discussed in Section 2. Hence, in practice, the recommended methodology is relatively close to that for many other evaluations even if less ambitious than either ‘overall reform programme’ social cost benefit based counterfactuals or formal control group random assignment comparators.

The logic of the recommended approach is as follows:

- (i) Establish the quality of governance of the regulatory entity as well as its staff resources, competence, etc;
- (ii) Establish how the regulator entity operates relative to regulated companies, Ministries, the legislature and legislative committees, consumer representatives, etc;
- (iii) Identify key regulatory issues and decisions;
- (iv) Identify key outcome indicators for the regulated industry (growth in number of connections, cost, and efficiency trends, investment performance, service interruptions, prices and profitability);
- (v) Estimate the role of the regulator in helping or hindering the achievement of the outcomes in the light of other factors, as well as how, and why this was so; and
- (vi) Derive lessons as to how performance can be improved in the future.

4.3 The Handbook Methodology and Evaluation Tools

The recommended approach’s main steps are, firstly, to establish governance characteristics and resources; secondly, to carry out an ‘analytic description’ of operations and outcomes; and finally, to relate regulatory activities and decisions to outcomes.

4.3.1 Evaluation of Regulatory Governance

On governance, the recommended method of analysis was relatively standard but it is much better developed and spelled out than previously. In particular, a hierarchy was presented of three ‘Meta-Principles’, ten ‘Principles’, and fifteen ‘Critical Standards’ (which were broken down into over 100 separate items).

Of these, only the Meta-Principles were deemed to be applicable to all regulatory entities. The Principles and Critical Standards are intended only for regulatory entities that go some way towards being designed and operated as ‘independent’ regulatory agencies. The Meta-Principles are intended to be non-controversial and applicable worldwide but it would be absurd to pretend that they are value-neutral. In particular, they assume some degree of both commercialization and of effective legal support for commercial operation.

The three Meta-Principles were specifically intended to apply to all infrastructure regulatory entities in all countries:

Without prejudging the institutional forms that these other regulatory governance systems may take, it is clear that a regulatory system can be effective only if it satisfies three basic meta- or higher-order principles:

- *Meta-Principle 1: Credibility*—Investors must have confidence that the regulatory system will honor its commitments
- *Meta-Principle 2: Legitimacy*—Consumers must be convinced that the regulatory system will protect them from the exercise of monopoly power, whether through high prices, poor service, or both.
- *Meta-Principle 3: Transparency*—The regulatory system must operate transparently so that investors and consumers ‘know the terms of the deal’ [or “rules of the game”]

(Brown et. al., 2006: 55).

The Ten Key Principles are relatively standard and similar to other specifications of regulatory principles for regulatory entities (e.g. Warrick Smith, 1997; Stern and Holder, 1999; UK Better Regulation Executive, 2001). They are:

- 1) Independence
- 2) Accountability
- 3) Transparency and Public Participation
- 4) Predictability
- 5) Clarity of Roles
- 6) Completeness and Clarity in Rules
- 7) Proportionality in Application
- 8) Requisite Powers
- 9) Appropriate Institutional Characteristics
- 10) Integrity of Conduct

The *Critical Standards* are a new departure and are designed to specify the appropriate mechanisms by which the Principles can be achieved in practice by regulatory agencies which have some degree of regulatory ‘independence’. They are:

- (i) Legal Framework
- (ii) Legal Powers
- (iii) Property and Contract Rights
- (iv) Clarity of Roles in Regulation and Policy
- (v) Clarity and Comprehension of Regulatory Decisions
- (vi) Predictability and Flexibility
- (vii) Consumer Rights
- (viii) Proportionality
- (ix) Financing of Regulatory Agencies
- (x) Regulatory Independence
- (xi) Regulatory Accountability
- (xii) Regulatory Processes and Transparency
- (xiii) Public Participation
- (xiv) Appellate Review of Regulatory Decisions
- (xv) Ethics

This somewhat summary list presents specific items under each heading in a 59 page appendix. The latter provides the basis for as full an evaluation of regulatory governance performance as desired – and a much more operational basis than before.

One final point on the evaluation of regulatory governance is that very considerable emphasis is placed throughout the Handbook on evaluating what happens in practice and not just what is specified in the regulatory law or other legal documents. Incorporating regulatory practice has been a major development in the literature on regulatory governance and its impact in recent years, e.g. as in Montoya and Trillas (2009) who find that adding information on whether regulatory commissioners serve out their terms of office is important as well as legal quality for explaining telecom penetration rates in Latin America and the Caribbean.

4.3.2 *Levels of Evaluation*

The Handbook proposed three levels of evaluation: ‘Short, Basic’; ‘Mid-Level’ (the main focus of the Handbook); and ‘In-Depth’. The first is intended as a diagnostic check on the basic characteristics of the sector via a structured questionnaire. This has good coverage on regulatory governance issues and industry structure, but relatively little on outcomes – primarily a few open-ended questions on successes and setbacks. The Handbook provides a suggested Questionnaire (in Appendix C).

The mid-level evaluation is intended to probe rather harder on regulatory governance issues (legislative and in practice) but is the first level at which evaluation of outcomes can seriously be achieved. There are again appendices with questionnaires and interview guidelines.

The in-depth evaluation is essentially the mid-level evaluation but with much more probing and a wider remit. The basic evaluation is conceptually different from them, but was intended not least to help provide comparable data for many countries.

The extent of the outcome-related work (regulatory substance evaluation) increases as the depth of evaluation goes up. In practice, it is difficult to draw a line between the mid-level and in-depth evaluations other than by time and budget. The evaluation of the Jamaican OUR (Office of Utility Regulation) discussed in Section 4.4 was an example of a mid-level evaluation which

effectively became an in-depth evaluation, primarily because of the logic of the enquiry and the requirements of the Jamaican authorities.

4.3.3 *Evaluations of Regulatory Substance – Regulatory Decisions and Sector Outcomes*

Regulatory agencies achieve their objectives by making decisions and these decisions are taken within the context of government policy. These may be stated or unstated, mutually consistent or not. Governments typically have a range of policy objectives for infrastructure industries. In the UK and other EU countries, we find targets for energy industry emissions, fuel diversity in generation, and renewable energy use, broadband penetration, railway use within total transport, etc. In developing countries, we find targets for rural electrification, piped water, and sewerage availability, etc. In many countries, we find targets for quality (e.g. supply interruptions), investment, prices, affordability and subsidies (including cross-subsidies).

Where government targets are explicitly stated and published, regulators should try to achieve them – *provided that* they are not in conflict with other legal obligations that they may have towards consumers and investors. But, governments often have multiple objectives and they may conflict with one another. In those circumstances, a major function of the regulator is to identify potential conflicts and to discuss with governments how they might be resolved.

The most common problems are where governments have policy targets which require substantial industry investment but the government is unwilling to allow prices to be raised to cover the expenditure. In some cases, this may be resolved by specific subsidies (e.g. railway subsidies and rural electrification subsidies). However, the most difficult problem is where governments have implicit, unpublished, or vague policy objectives.

Contrary to some suggestions, the regulator's task is always easier the more clearly and fully governments state their policy objectives. Whether and how far regulators further those objectives is an important issue in evaluating the performance of regulatory agencies. Hence, regulatory decisions refer to any action or inaction by the regulator that affects the interests of participants in the regulated sector – consumers, producers, investors, or governments.

The objectives of a good regulatory system are:

- It produces a flow of good regulatory decisions.
- It minimizes the number of poor or mistaken decisions.
- It speedily corrects errors.
- It does not repeat mistakes or poor decisions.
- It learns from regulatory good practice in other sectors and countries.

Failing to meet these objectives implies significant evidence of flaws in the design and operation of the regulatory system.

If the true test of regulatory effectiveness is the impact on sector outcomes, the first issue is what those outcomes should be. The set chosen relate to *commercialized* infrastructure industries and are set out below.

It should be noted that the whole framework of the evaluation of regulatory substance set out in the Handbook applies *only* to commercialized or commercializing companies. These would primarily be regulated enterprises with significant amounts of private involvement (or at least private financing of investment) in some or all elements. However, it could also include publicly owned enterprises where investment was financed on market terms and the government required the enterprise to earn a normal rate of return on capital (Brown et. al., 2006: 88-91; Stern, 2007).

The Handbook identifies eight headings for industry outcomes. These headings are¹⁷:

- 1) Output and Consumption
- 2) Efficiency
- 3) Quality of Supply
- 4) Financial Performance
- 5) Capacity, Investment and Maintenance
- 6) Prices
- 7) Competition
- 8) Social Indicators

Some of the variables listed are *final* outputs of the industry (e.g. consumption levels, quality of supply). Most of the other variables listed are *intermediate* outputs (e.g. efficiency, costs and prices, and competition indicators). However, investment and financial performance have elements both of final and intermediate output because of their impact, firstly, on the sustainability of output and consumption levels; and, secondly, on other aspects of economic management (e.g. the government's fiscal position and inflation).

A fundamental evaluation issue is how to isolate the impact of regulatory decisions from those of other major factors. This means that the evaluator must be aware of and take account of all relevant factors besides regulatory decisions. This is because these other factors often have a rather larger impact on industry outcomes than regulatory decisions. Examples of these other major factors include, among others:

- The appropriateness and coherence of the chosen industry and market structure (or lack of coherence, as in the 1990s electricity reform structures adopted in California and the Ukraine).
- Problems arising from inconsistencies in government policy and/or government unwillingness to allow the regulatory agency to carry out its functions (e.g. Russia and India, at least until relatively recently).
- External pressures, particularly the impact of macroeconomic and exchange rate crises on costs and prices (e.g. Argentina and South East Asia in the late 1990s).

The evaluator should draw attention to these factors and how the regulatory system responded to the difficulties. Did the regulator assist or hinder responses to them? More importantly, was the regulator allowed to respond? In the case of the exchange rate crises identified above, the response of the governments above was not just to suspend effective regulatory decision making but also not to use the regulator to help achieve sustainable debt 'workouts'.¹⁸

This methodology allows a *marginal* evaluation approach to the impact of regulatory agencies as well as the analysis of the impact of specific decisions, and this can be related back to the characteristics of the agency (its governance, its resources, its procedures, etc). It also enables the identification of 'good' or 'bad' decisions in terms of objective information on industry/sector outcomes. This includes information from interviews with a range of involved players and the collection, and analysis of statistical data.

For transitional or intermediate regulators (i.e, infrastructure regulators in countries unable and/or unwilling to establish an 'independent' regulator), the evaluation task is the same, except that in addition, the evaluation needs to:

- Establish whether and how far the observed regulatory system has moved to improve its regulatory practices relative to previous arrangements;

- Assess the regulatory arrangements and industry outcomes with those in comparable countries;
- Identify the internal incentives and pressures that are:
 - likely to improve regulatory processes and industry outcomes
 - likely to impede or retard progress, or threaten the viability of current regulatory arrangements, or both.

4.4 Application of the Handbook Methodology to the Evaluation of the Performance of the Jamaican Office of Utility Regulation

The first known commissioned review of an infrastructure regulatory agency using the Handbook methodology was the 2006 PPIAF/World Bank commissioned Regulatory Impact Assessment¹⁹ of the Office of Utility Regulation (OUR) in Jamaica.²⁰ This is a multi-sector regulator which is responsible for the economic regulation of telecommunications, electricity, water and some public transport facilities.

The evaluative purpose of the study was, firstly, to assess the efficiency and effectiveness of the OUR (and other agencies in Jamaica) with a view of establishing whether its cost represented ‘value for money’; and, secondly, to do a strategic review of the OUR’s past performance and examine choices for the future. In particular, there was the question as to whether telecoms should be carved out of the OUR and put together with radio spectrum and broadcast content regulation in a Jamaican communications regulator (a Jamaican Ofcom/FCC). Finally, there was the issue of what aspects of regulation of the different industries had gone well or badly, and the role of the OUR in that so as to learn lessons for future activities.

The content of the evaluation included material on the OUR budget and efficiency trends, and a discussion of regulatory governance issues and legal gaps. However, the main novelty relative to previous evaluations of regulatory agencies was the identification of key decisions and the assessment of their impact for the regulated industries following the World Bank Handbook approach as set out in section 4.3.3 above. Focusing on telecommunications and electricity, to estimate the impact of the OUR on outcomes in these industries, an ‘Event Analysis’ was undertaken which examined key regulatory decisions. This required information on:

- the history of agency activities and decisions over a number of years (at least five, preferably more); and
- an ability to relate key decisions to demonstrable industry outcomes, taking account of implementation and behavioural response lags.

In fact, typically information was available covering 10 years or more.

The message to be drawn from the evaluation of the OUR, however, is that it is enormously difficult to separate evaluations of regulatory practice from evaluations of the overall governmental policy within which a regulatory system is placed. In the Jamaican context it is arguable that an appraisal of regulation should focus not on the question whether privatizing and liberalizing measures have delivered desired outcomes but on whether, given those measures, the regulators have done a good job. Defining which policies and decisions constitute part of the policy framework, and which should be seen as integral to the regulatory regime is, however, a considerable challenge. Nevertheless, especially when compared to Trinidad & Tobago, the

development in telecommunications pointed to a broadly positive impact of the initial liberalization choice (by the government) and of the OUR's implementation decisions.

With regard to Jamaican electricity, it was found that the privatisation strategy had been seriously flawed. An Event Analysis revealed that, taking full account of subsidies, non-fuel generation costs in Trinidad and Tobago in 2002 were around 4 US cents/kWh as against almost 8 cents/kWh in Jamaica. The regulatory evaluation question, however, was whether or not the OUR could have done more to improve matters within the privatisation legislation and its own powers. Looking at the legislation and the licence, there seemed to be no reason why the OUR could not have pursued separate accounts and separated generation, and network tariffs. Even for the areas where licence and/or legislative changes would have been necessary (e.g. separate businesses and removing generation planning and dispatch from JPSCo), it would have been possible for the OUR to pursue the changes energetically both in private and public. They had done none of these and the evaluation provided a potential agenda to the OUR, and the government for trying to rectify the problems in Jamaican electricity.

4.5 Application of the Handbook-style Methodology to other Regulatory Issues

Stern (2007) conjectures that the general framework proposed in the Handbook may well be relevant to other types of regulation e.g. financial regulation, health and safety regulation, consumer protection regulation, etc. In particular, the focus (a) on key decisions and (b) on consumer and industry outcomes seems very general. Hence, the type of outcome evaluation discussed above may well be a useful analytic component of a potentially wide range of ex post regulatory evaluation. Indeed, the UK Financial Services Authority appears to be taking an approach of this type, particularly over the regulation of retail financial products (see Oxera, 2008). More obviously, there has been a considerable evaluation work on the merger decisions of competition authorities and whether or not the required remedies to eliminate potentially adverse effects have been justified and effective. The UK Competition Commission has carried out and in 2008 published the results of such a study. Similar studies have been done in the EU (the 2005 Merger Remedies Study) and the US. A good recent survey of the latter is by Carlton (2007) whose main conclusion is that effective merger policy evaluation requires data on the predictions of the government competition agencies on post-merger markets as well as data on the relevant market pre and post-merger.

It is no coincidence that the areas of financial regulation and merger decisions have become the focus of ex post evaluation. Both are areas where policy has been controversial. DG Comp had previously had several cases where their decision to prevent mergers has been overruled by the Court of First Instance (*Airtours*, *Schneider*, and *Tetra Laval* cases) so that there was a political and legal imperative on them to establish new merger guidelines. The EU Merger Remedies evaluation study was part of that process. Similarly, the published UK Competition Commission evaluation explicitly refers to its obligation to consider the effectiveness of its decision making since it - and not the relevant Minister - has since 2003 had the sole responsibility both for the choice of remedies and for making them work. Hence, the Competition Commission has carried out the evaluation so that it can learn lessons from its previous decisions. Methodologically, of the studies above, most rely primarily on interview material and simple statistical comparisons but some of the merger evaluation research has ventured into highly technical econometric analysis.

5 CONCLUSION

The crucial issues for effective ex post evaluation are the creation of a robust counterfactual to the change and the associated topic of establishing the reasons for differences between outturns and the counterfactual. Both of these were important in the evaluation of the Jamaican OUR discussed in Section 4.4. These issues are far from easy for public expenditure decisions and programmes but they seem to be a lot more difficult for regulatory decisions and especially, for the evaluation of the performance of regulatory agencies. In particular, the random assignment control group model that is again highly fashionable as a way of handling the difficulties of constructing a robust counterfactual is totally unworkable in the regulatory context as regulatory decisions – and certainly the creation of a regulatory agency – is anything but random.

A major reason for the extra difficulties of evaluating regulatory decisions are that they are typically only one determinant of the outcomes of the regulated industry and typically often a relatively minor aspect. This was shown clearly in the evaluation of the Jamaican OUR and its regulatory performance with regard to Jamaican electricity. It had very little leverage and had done little to stretch, or enhance what limited powers it had.

When it comes to evaluating the performance of regulatory agencies, the main difficulty lies in identifying the target of evaluation – be that a government policy or a regulatory regime designed to implement that policy. A further problem is that the counterfactual of no agency asks far too much. The evaluator is not considering a marginal change; it involves having to consider what set of structural choices and decisions might have been taken by some alternative, hypothetical regulatory framework. As this is unworkable, the ex post evaluation focus turns to evaluating sets of decisions taken by the regulator, relative to other decisions that *it* might have taken. Those decisions have to be big enough significantly to affect consumer, efficiency, and investment outcomes but not so large as to recast the sector. Finding enough such decisions can be difficult.

In practice, case-study based evaluations typically use a mixture of pre and post decisions and comparisons with the results of comparable events in similar countries. Hence, they require a lot of data and a great deal of contextual information with which to interpret the data. They are an art, at least, as much as a science – and maybe rather more so.

In practice, ex post evaluations are always considerably strengthened by the existence of an ex ante option appraisal, such as a regulatory impact assessment or similar. This is suggested by the ROAMEF cycle where ex ante appraisals are recommended and ex post evaluations carried out where useful and appropriate. The existence of an ex ante appraisal has been shown to be crucial for evaluations of merger decisions and remedies in competition policy cases.

The discussion above focuses on other than formal statistical models of evaluation. Over the last 10-20 years, there has been major growth in the use of econometric techniques in this area, particularly with the development of many more panel data sets. Panel data econometrics has been used considerably both for evaluating the effectiveness of infrastructure regulators (primarily, but not exclusively) in developing countries and of merger decisions and remedies, particularly in the US.

These methods are a useful addition, not least because they produce results from which one can generalise as well as results that can be subject to repeat testing. However, they cannot be guaranteed to show causation rather than just correlation and, crucially, they do not explain *how*

and why positive or negative results occur. When one adds in the fact that they are not readily explicable to non-statistically minded people, it can be seen that they do not provide the basis for learning what the best case study designs can do. They complement analytical case studies but are not a substitute for them.

In consequence, whatever the difficulties, it is essential that there is continuous development and improvement of analytical case study approaches like that of the World Bank Handbook on evaluating infrastructure regulatory frameworks. This can and should be accompanied by development of the formal models, as the two methods (formal modelling and case studies) provide mutually supportive insights and examples. All of the above considers the technical aspects of evaluating regulatory entities. But, evaluation is much more than just a technical issue. There are major political economy implications and, indeed, fundamental political choice issues.

Evaluation is primarily about accountability. In the UK, the institutional auditing of Exchequer payments goes back to 1559 and there is a reference in a document from 1314 to an 'Auditor of the Exchequer'. The modern development of expenditure audit for use in public expenditure control was a 19th century phenomenon clearly linked to improving the accountability of the executive to the legislature; and, with the growth of democracy and the mass media, to the wider public. 'Value for money' audit is part of this and it has traditionally examined how well expenditure has been spent as well as whether it has been spent legitimately; this has led to the rise of 'policy audit'. The same arguments have now been applied to regulation and have become progressively more vocal with the increasing role of regulation. We now even have strong lobbies for 'regulatory budgets'

In consequence, the growth of regulatory evaluations is part of the general demand for greater government accountability and the need to monitor performance. This demand is widespread in many countries and in single party as well as multi-party states. The pressure for effective evaluation comes both from technocratic pressures e.g. from ministries of finance (particularly in times when there are strong pressures on the amount of affordable public expenditure) and from general political pressures. The latter is more evident with regulation, as seen in the attempt of Britain and some other EU member states to obtain better justification for EU regulatory initiatives.

Given these wider pressures, it is no surprise that the pressures for effective evaluation are strongest on aid agencies in general and, in particular, on multilateral aid agencies like the World Bank. They have a wide range of countries from whom they obtain grants, borrow and to whom they lend at highly concessionary as well as near commercial rates. Many of the borrowers and lenders are important and politically powerful countries. This is a major reason why evaluation of regulation and regulatory agencies is an area in which multilateral institutions often lead in the development of new methods as well as in their dissemination.

As a consequence of these pressures, we expect regulatory evaluation to grow and become more important in coming years. The pressures for effective ex post evaluation should also lead to improvements in the techniques and methods of evaluation – as well as the discussion of the results of such evaluations. The debate on how and why financial regulation was deficient pre-2007 and the use of evaluation of past financial regulation frameworks and rules for devising new methods well demonstrates the potential role of ex post evaluation for regulatory entities of all types.

ANNEX 1: INFRASTRUCTURE INDUSTRY OUTCOMES FOR EVALUATION PURPOSES

Regulatory decisions help affect electricity industry performance on the following outcomes:

1) Output and Consumption

Household and business access levels
Consumption levels and growth rates per head and per unit of GDP
Levels of unsatisfied demand
Emission and pollution levels

2) Efficiency

Productivity levels and growth rates
Cost levels and changes
Capacity availability & utilisation; losses (technical and commercial)

3) Quality of Supply

Continuity of supply
Quality of supply and customer service

4) Financial Performance

Financial surpluses and losses, achieved rates of return
Measures of indebtedness and interest burden

5) Capacity, Investment and Maintenance

Capacity levels and margins
Levels of investment and share of private and foreign investment
Levels of maintenance expenditure

7) Prices

Relationship of prices to full economic costs (including a reasonable rate of return on assets)
Explicitness, transparency and efficiency of subsidies and cross-subsidies
Tariff design that promotes technical and economic efficiency in production, fuel use and consumption)
Degree to which environmental costs included in economic costs and prices

8) Competition

Well-functioning bid auction markets for concessions and IPP contracts with a sufficient number of bidders

Well functioning generation and supply competition markets

(Equivalent indicators can readily be constructed for infrastructure industries other than electricity)

9) Social Indicators

Affordability of supply – particularly for low income consumers

Impacts on economic development

Three points arise:

- (i) This list is, of course, not exhaustive and can be extensively elaborated depending on the depth and detail of evaluation required.
- (ii) The list above has been devised as reflecting the major issues for electricity/energy evaluation. It can readily and straightforwardly be amended for other infrastructure industries; indeed, most of the individual entries would stay the same.
- (iii) Some of the variables in the list above are *final* outputs of the industry (e.g. consumption levels, quality of supply, impacts on economic development). Most of the other variables listed are *intermediate* outputs (e.g. efficiency, costs and prices, competition indicators). Investment and financial performance have elements both of final and intermediate output. This is because of their impact, firstly, on the sustainability of output and consumption levels; and, secondly, on other aspects of economic management (e.g. the government's fiscal position and inflation).

REFERENCES

- Andres, L., Guasch, J. L., and Azumendi, S. L. (2008). 'Regulatory Governance and Sector Performance: Methodology and Evaluation for Electricity Distribution in Latin America', *World Bank Policy Research Working Paper No.4494*.
- Banerjee, A. and Duflo, E. (2008). 'The Experimental Approach to Development Economics', *MIT Mimeo*, Sept 08. Available at: <http://econ-www.mit.edu/files/3159>.
- Besley, T. (2002). 'Elected versus Appointed Regulators', *LSE Mimeo*.
- Better Regulation Task Force (2001). *Economic Regulators*, London, HMSO.
- Brown, A., Stern, J., and Tenebaum, B. (2006). 'Evaluating Infrastructure Regulatory Systems', Washington: The World Bank.
- Carlton, D. W. (2007). 'The Need to Measure the Effect of Merger Policy and How to Do It', *Economic Analysis Group, EAG Working Paper No. 07-15*
- Competition Commission (2008). *Understanding Past Merger Remedies: Report on Case Study Research*.
- Cubbin, J. S. and Stern, J. (2006). 'The Impact of Regulatory Governance and Privatization on Electricity Generation Capacity in Developing Countries', *World Bank Economic Review*, pp 115-141.
- Deaton, A. (2009). 'Instruments of Development: Randomization in the Tropics and the Search for the Elusive Keys to Economic Development', *Keynes Lecture 2008*.
- DG COMP European Commission (2005). *Merger Remedies Study*, Public Version.
- Edwards, G. E. and Waverman, L. (2006). 'The Effects of Public Ownership and Regulatory Independence on Regulatory Outcomes: A Study of Interconnect Rates in EU Telecommunications', *Journal of Regulatory Economics, Vol. 29, No.1*.
- Ehrhardt, D., Groom, E., Halpern, J. and O'Connor, S. (2007). 'Economic Regulation of Urban Water and Sanitation Services: Some Practical Lessons', *World Bank Water Sector Board Discussion Paper Series Paper No. 9*.
- Foster, C. D. and Beesley, M. E. (1963). 'Estimating the Social Benefit of Constructing an Underground Railway in London', *Journal of the Royal Statistical Society (Series A)*.
- Gasmi, F. and Virto, L. R. (2008). 'The Determinants and Impact of Telecommunication Reforms in Developing Countries', *IDEI Working Paper No. 530*.
- Guasch, J.L. and Straub, S. (2006). 'Renegotiation of Infrastructure Concessions', *Annals of Public and Cooperative Economics*, 77 (4): 479-494.
- Gutierrez, L.H. (2003). The Effect of Endogenous Regulation on Telecommunications Expansion and Efficiency in Latin America, *Journal of Regulatory Economics*, 23 (3): 257-286.
- Hahn, R. (2008). 'Designing Smarter Regulation with Improved Benefit-Cost Analysis', *Reg-Markets Center Working Paper 08-20*
- Heckman, J. J. (1992). 'Randomisation and Social Policy Evaluation', in *Evaluating Welfare and Training Programmes*, C. Manski and I. Garfinkel (eds.), Cambridge, MA: Harvard University Press.
- Horton, G. R. and Littlechild, S. (2002). 'Re-evaluating the Costs and Benefits of UK Electricity Privatisation', *Mimeo*.
- H.M. Treasury (2003). 'The Green Book'. Available at: <http://greenbook.treasury.gov.uk/>.
- Layard, P.R.G. (1972). 'Cost-Benefit Analysis', Penguin, London (2nd Revised Edition with S. Glaister, CUP, 1994).
- Levy, B. and Spiller, P. (1996). *Regulation, Institutions and Commitment*, Cambridge, UK: Cambridge University Press.

- Little, I. M. D. and Mirrlees, J. (1974), *'Project Appraisal and Planning in Developing Countries'*, London: Heinemann Educational.
- Maiorano, F. and Stern, J. (2007). 'Institutions and Telecommunications Infrastructure in Low and Middle-Income Countries: The Case of Mobile Telephony', *Utilities Policy*, 165-182.
- Montoya, M.A. and Trillas, F. (2009). 'The Measurement of Regulatory Independence in Practice: Latin America and the Caribbean', *International Journal of Public Policy* (Forthcoming).
- National Audit Office (2006). 'Evaluation of Regulatory Impact Assessments 2005-06', Available at: http://www.nao.org.uk/publications/nao_reports/05-06/05061305.pdf.
- (2007). 'A Review of Economic Regulator's Regulatory Impact Assessments'. Available at: http://www.nao.org.uk/publications/nao_reports/07-08/economic_regulators_impact_assessments.pdf
- Newbery, D. and Pollitt, M. (1997). 'The Restructuring and Privatisation of the CEGB – Was it Worth It', *Journal of Industrial Economics*, 45(3): 269-303.
- Ofcom (2006). 'Television Advertising of Food and Drink Products to Children – Options for New Restrictions'. Available at: <http://www.ofcom.org.uk/consult/condocs/foodads/>.
- Oxera (2008). 'Towards Evaluating Consumer Products in the Retail Investment Products Market: A Methodology Prepared for the Financial Services Authority', Financial Services Authority.
- Ravallion, M. (2008). 'Evaluation in the Practice of Development', *World Bank Policy Research Working Paper No. 4547*.
- Rodrik, D. (2008). 'The New Development Economics: We Shall Experiment, but How Shall we Learn?'. Available at: <http://ksghome.harvard.edu/~drodrik/The%20New%20Development%20Economics.pdf>.
- Shirley, M. (ed.) (2002). *Thirsting for Efficiency: The Economics and Politics of Urban Water System Reform*, Washington, D.C.: The World Bank.
- Smith, W (1997). 'Utility Regulators – The Independence Debate', *World Bank Public Policy for the Private Sector Note No. 127*.
- Stern, J. (1997). 'What Makes an Independent Regulator Independent?', *Business Strategy Review*, 8(2): 67-74.
- Stern, J. and Holder, S. (1999). 'Regulatory Governance: Criteria for Assessing the Performance of Regulatory Systems: An Application to Infrastructure Industries in the Developing Countries of Asia', *Utilities Policy*, 8(1): 33-50.
- Stern, J. (2007). 'Evaluating Infrastructure Regulators: Developing UK and International Practice', *CRI Occasional Lecture, No. 17*, January 2007.
- Wallsten, S. (2002). 'Does Sequencing Matter? Regulation & Privatisation in Telecommunication Reform', *World Bank Working Paper No. 2817*.

END NOTES

-
- ¹ Cost effectiveness analysis may be used as a fallback e.g. if the topic of concern is alternative methods of achieving an outcome where the benefits are fixed or common across all alternatives.
- ² *Green Book: Appraisal and Evaluation in Central Government*, H.M. Treasury, 2003. Available at: <http://greenbook.treasury.gov.uk/>. It is worth noting that the first four editions of the Green Book were published by the Treasury and only the 5th and subsequent editions have been published by the (Government) Stationery Office. The latest edition, although published by the Stationery Office, while clearly recognising the contributions and comments of other government departments, is still described on its cover page as ‘Treasury Guidance’.
- ³ For financial regulation, the claims of excessive regulation have gone sharply into reverse since the onset of the financial crisis in 2008. Financial regulation, particularly bank regulation, is much more complex than infrastructure or consumer protection legislation as it involves difficult systemic and financial market issues, as well as, protection of retail finance buyers, and sellers (see Moloney in this volume).
- ⁴ *Green Book: Appraisal and Evaluation in Central Government* (2003). Available at: <http://greenbook.treasury.gov.uk/>.
- ⁵ See, <http://www.berr.gov.uk/whatwedo/bre/policy/scrutinising-new-regulations/page44076.html>
- ⁶ *Evaluation of Regulatory Impact Assessments 2005-06*, NAO, 2006. Available at: http://www.nao.org.uk/publications/nao_reports/05-06/05061305.pdf.
- ⁷ *A Review of Economic Regulator’s Regulatory Impact Assessments*, NAO, 2007. Available at: http://www.nao.org.uk/publications/nao_reports/07-08/economic_regulators_impact_assessments.pdf.
- ⁸ See, <http://www.ofcom.org.uk/consult/condocs/foodads/>.
- ⁹ See, http://www.nao.org.uk/ria/ria_our_work.htm for a listing of NAO regulatory investigations.
- ¹⁰ HM Treasury Green Book (2006: 46).
- ¹¹ The relatively optimistic view of regulators can be over-emphasised and it would be strongly denied by many, particularly by followers of Buchanan-style public choice theory.
- ¹² Of course, these can be *either* complements to *or* substitutes for research funded, independent, not-commissioned evaluations. That depends on how and why the commissioned researchers are chosen.
- ¹³ For further examples of single country infrastructure industry case-study evaluations, see Brown et al (2006: 33) and Appendix G. The World Bank website and the PPIAF website have many others.
- ¹⁴ In technical language, there is a serious identification problem.
- ¹⁵ Unobservable fixed effects are estimated via fixed effects or similar econometric techniques. They should capture important elements of variations in country governance and its reputation – in practice as well as on paper.
- ¹⁶ More recently, Gasmi and Recuero Virto (2008) explore endogeneity issues in more detail for telecoms, taking account of mobile-fixed line interactions. They do not find the positive effects of regulation as in their 2006 paper. It remains to be seen whether these results are replicated in further work.
- ¹⁷ The list above is a summary list. A more fully elaborated list with sub-headings under each entry is set out in the Annex of this chapter.
- ¹⁸ In Argentina and other Latin American and Asian countries, infrastructure industry investment is often financed by debt denominated in foreign currency but with services sold in home currency prices. Following a major depreciation of the home currency, these debt contracts become unviable and need major renegotiation, or replacement. Such renegotiations are known as a ‘debt workout’ process.
- ¹⁹ This RIA was an ex post evaluation, not (as in standard UK terminology) an ex ante appraisal.
- ²⁰ The study was carried out by CEPA (Cambridge Economic Policy Associates). The study has not been formally published but it has been referred to in public documents. See, for instance pages 5 and 10 of the 2006-2007 Annual Report of the Jamaican Office of the Cabinet, http://www.cabinet.gov.jm/files/u2/Annual_Report_2006-2007.pdf?phpMyAdmin=36964530831c7b5cd24342ae2600c405.