British infrastructure policy and the gradual return of the state

Dieter Helm*

Abstract The state of British infrastructure is the source of almost continuous study and criticism. There is now a host of new demands in water, energy, transport and communications which require a step change in investment. The Coalition government has made a series of reforms—to planning, finance, and the setting of priorities—which have been drawn together in National Infrastructure Plan statements. A number of high-profile and large-scale projects have been advanced. This paper sets out the evolution of these policy interventions and explains why government involvement is critical to their success. It documents the gradual return of the state, as part of a process of underpinning the investment costs, in the provision of credible contracts, and in closing the gap between public and private costs of capital. It suggests that rather than pursue infrastructure policy as a set of priority projects, greater use of regulated asset bases and attention to the time-inconsistency problem would better achieve the overarching objectives.

Keywords: infrastructure, regulation and industrial policy, investment, infrastructures, other public investment, capital stock

JEL classification: O18, L52, D92, H54

I. Introduction

The state of British infrastructure is a source of almost continuous study and criticism. Congested roads, expensive and crowded trains, polluting and expensive energy, and slow and limited access to broadband are repeatedly identified as ‘problems’ which governments ought to sort out. These criticisms are not new—the British have been criticizing their infrastructure for at least a century. Nor are they confined to Britain. But they have taken on a new urgency as the demands placed upon them have changed, sometimes dramatically, and their role in supporting economic growth and recovery has become a priority.

These new demands come in a variety of shapes and forms. In water, European directives have necessitated a clean-up of rivers and beaches, and a combination of weather events and building on flood plains has highlighted the inadequacy of flood defences.

* New College, Oxford, e-mail: dieter@dhelm.co.uk

Comments by Cameron Hepburn have been greatly appreciated. Any errors, however, remain mine.

1 See President Obama’s State of the Union Speech (Obama, 2013).
In energy, a combination of the ageing of existing power stations—built in the context of the industrial demand of the 1960s and 1970s—more European directives, and the decarbonization agenda require a wholesale replacement of much of the existing capacity and, indeed, the electricity systems themselves. New smart technologies will bring added opportunities and pressures. The decline of North Sea gas reserves means new gas import pipelines and liquefied natural gas (LNG) facilities are required. Roads and railways require expansion, and high-speed rail requires a whole new network. Airport capacity in the south-east is inadequate, and Britain’s early lead in telecommunications has given way to a number of urgent and serious problems in broadband coverage and speeds. The combined total capital expenditure is of an order of magnitude greater than current investment.

To address these demands, successive governments have recognized they need to ‘do something’. Often this has been to set aspirations and targets, rather than do much about the underlying incentives to invest. Specific projects have been championed—upgrading the London Underground, Crossrail, the Thames Tideway, the building of offshore wind farms, and High Speed Rail 2 (HS2). The Coalition government introduced planning reforms and a National Infrastructure Plan, and the Labour opposition has established the Armitt Commission to bring forward its own infrastructure plans and policies.

There has been no lack of activity. Yet despite all these initiatives, progress has been slow at best. Part of this can be put down to the credit crunch and the economic circumstances, but the gap between aspiration and outcomes suggests there may be deeper forces at work.

The structure of this article is as follows. Section II explains why government involvement is critical, and why infrastructure provision is likely to be sub-optimal in the absence of intervention. Section III describes the evolution of infrastructure policy and the initiatives announced since 2010. Section IV sets out the required reforms if the infrastructure agenda is to be addressed. Section V considers the return of direct state involvement and section VI concludes.

II. The underlying market failures—why intervention is essential

The provision of infrastructure almost always involves some form of state intervention. The explanation lies in the underlying multiple market failures. Infrastructure typically has high fixed and sunk costs at the outset, and then low or very low marginal costs, and the assets tend to be long lived. These characteristics interact. The investor needs to recoup these sunk and fixed costs over the lifetime of the project, but once the project is completed the efficient price is equal to the marginal, not the average cost, and the infrastructure will be worth operating as long as the marginal costs can be recovered with some contribution to the fixed costs. The investor is therefore exposed to the time-inconsistency problem: to being promised that the full costs will be recovered ex ante, and then being squeezed back down to marginal costs ex post. The long life of the project adds a further twist, since the longer the time for cost recovery, the more chance that there might be technical progress that strands some or all of the asset.
What the investor therefore needs is a credible contract: a guarantee that in return for sinking the upfront costs, these will be recovered. The only serious party to such a contract for large-scale national infrastructure is the state, and such contracts can be mediated through a variety of institutions within the framework of the state’s protection and its guarantees. As we shall see, the private finance initiative (PFI), public–private partnerships (PPP), and the regulated asset base (RAB) in utilities are just some of the various possible ‘solutions’ to this contracting problem. Where they prove difficult to provide, state ownership and state procurement have often been deployed.

This contracting problem is not always apparent, and it is not the only market failure. In the early days of new infrastructures—for example, in early train lines, electricity networks, and now broadband—investors can often charge high prices and exploit temporary monopolies. There can be investment races to dominate markets, with the hope of being able to exploit the subsequent dominance. Often, too, this is a period of irrational exuberance. Big prizes in terms of profits for the few can act a bit like lottery tickets: irrational investors stake too much on a small probability of a large pay-out, with the result that there can be significant over-investment and subsequent creative destruction. Such exuberance can be large enough to destabilize the whole economy—as with the railways and electricity and, more recently, the communications and dotcom bubbles. Once the initial burst is over, the new infrastructures tend to oligopolize and then the state intervenes to guarantee supplies and regulate prices. That is what happened to trains and electricity, and may now be happening to broadband.

The other market failures are many and vary greatly from case to case. Most infrastructures end up providing what come to be regarded as social primary goods and services—things that people need to participate in society and to protect some minimum threshold standard of living. These requirements vary over time, so that although there may be an absolute component, the infrastructures that meet this may vary with the social context and the shape of the economy and society. Broadband is now becoming necessary to participate in society, and it is joining the earlier list added by technical progress, including, notably, electricity.

Infrastructure can have a variety of positive and negative associated externalities. The most obvious example is the relation between the energy infrastructures and carbon emissions, but almost all have external effects. Much pollution goes through infrastructure networks—in addition to energy, transport and sewerage systems have significant environmental externalities.

Taking these multiple market failures together, the optimal level of infrastructure provision is very unlikely to be provided by the market without intervention. And since infrastructure is necessary not only for participation in society, but also for much economic activity, the costs of non-intervention are likely to be high. Yet recognizing that markets will not deliver the optimal level of infrastructure provision does not solve the problem of defining just what that level should be, nor ensure that government failures will not be commensurately large.

---

2 A very early example is provided by the development of postal services in the sixteenth century between England and the continent, and along the major roads. See Campbell-Smith (2011, ch. 1).
3 The 3G licences and the roll-out of cable are recent examples.
The conventional economic approach is to use some form of cost–benefit analysis (CBA) and, in particular, to use the kind of project-appraisal techniques that have become enshrined in government manuals—such as the Green Book (HM Treasury, 2003). Yet a moment’s reflection indicates how weak such techniques are when it comes to deciding how much infrastructure to provide. For infrastructure typically comes in systems, not discrete bits. Choosing what sort and level of infrastructure to supply is not a marginal decision. It is often about one system or another. Marginal analysis—as the core of CBA—has little obvious to offer.

An example illustrates this point. Take the HS2 proposal to build a high-speed rail link from London to Birmingham and then beyond to the north (DfT, 2013). The Department for Transport (DfT) has conducted a CBA (DfT, 2012) which shows that the benefits exceed the costs. However, the gap between the two has been challenged, and it is reasonable to argue that on a conventional basis they are actually quite close. Yet the question is not whether to build a new line to Birmingham from London, or to weigh up the costs and benefits to the existing economy. It is whether to have a high-speed rail system, within which the particular section would fit. From an analytical perspective, the costs and benefits of the system as a whole need to be taken into account first, and then this necessary initial part of such a network slots into place. In this context, it is important to recognize that a high-speed rail system aims to change the economy—as railways did in the nineteenth century, and electricity did in the twentieth century. The conventional assumption underlying CBA—that ceteris paribus does not hold.

How should the decision to choose a particular network system be made? This involves an economy-wide analysis. Building such a network will impact on total investment and hence economic output as a whole. It will alter locational decisions for households and companies. It will change cost functions. Thus the choice of system fits into a wider question: what sort of infrastructure as a whole is required to underpin the future economy? Some countries, such as China, do indeed take such a high-level approach, with Five Year Plans and state direction. European countries once planned their economies, too, in the post-war period up until the 1970s. The results in both China and earlier in Europe are open to serious challenge, and yet much of the infrastructure still around now is the product of such a national planning process. Examples include the electricity grids, the French nuclear programme, and, at a local level, the reservoirs, sewers, and water systems driven through by local government municipalities. Roads and rail are still planned in most countries.

National plans and government decisions feed through into a further major reason for government intervention. Decisions to build infrastructure usually involve access to and development of private and public property. The planning regime is a necessary filter for investment. The planning permission process tends to be set into national and local strategic development plans, in many countries reinforced by regional development agencies. Development requires infrastructure, and infrastructure provision tends to drive development. Having some sort of national, regional, and local planning process is the norm in most developed countries, and in many developing countries, too. Otherwise property rights would have limited content.

---

5 Notably by Glaister (2011).
It is hardly surprising, therefore, that infrastructure is set in the context of a contractual framework, with government responsible for taking system-wide decisions, and reinforced by some form of planning regime. The question is not whether the state should intervene, but how. The next step is to see how this worked out in the British context, and whether the current reforms are good solutions to these market failures.

III. The evolution of infrastructure policy

Britain’s approach to infrastructure has been through many phases. For the most part it has been concerned with two core issues: ownership and competition. After the Second World War, the answer was nationalization and monopoly. During the period 1980–2010, the answer was privatization and competition. The focus on these two dimensions has verged on obsession, with ideology often triumphing over analysis.

Public ownership gradually crept in. In the nineteenth century local government provision played a key role, notably for water and sewerage. Great city corporations in Manchester, Liverpool, and Birmingham planned and built reservoirs for fresh water supplies from the Lake District and Wales, and the sewers of London were a local government initiative. In the twentieth century local municipalities provided not just water, but also electricity and town gas, and became involved in bus services, other local transport services, and even regional airports.

This public-sector approach began to gather pace nationally in the inter-war years, notably with the development of the national electricity grid under the auspices of the Central Electricity Board (CEB), and, as total mobilization was necessitated by the demands of the Second World War, the state became the main economic actor. After the war, all the main infrastructures were controlled by the state, and the nationalized industry model became the main vehicle to develop and plan the new infrastructures. Indeed, these stretched beyond energy, post, and telecommunications to road and rail transport and into housing, health provision, and education. Competition and private provision were replaced by statutory monopoly and state ownership. It was a model replicated to a large extent across Europe.

This nationalized industry model worked well, in the sense that the infrastructures managed to cope with the great economic boom from the end of the Second World War until the 1970s. Motorways and power stations, some of them nuclear, were planned and delivered. But by the 1970s, the cracks were beginning to show, and in three different but related ways. First, the production monopolies had generated parallel labour monopolies, which in turn had created a national problem of strikes and other forms of industrial ‘action’ and, in the process, demonstrated how important infrastructures were to the functioning of the economy. Second, the investment and operating-cost performance came under scrutiny, and in a number of cases was deemed to be weak in comparison with the private sector.

But over and above these economic efficiency issues, the third problem with the nationalized model was the growth of the state as a share of the economy, and with it the levels of taxation, which had reached a limit necessitating public expenditure controls. For political reasons, and under budgetary pressure, it was capital rather than current spending that got cut. The result was that the investment that was needed to
maintain and enhance the infrastructure was no longer easily directly financed by the Treasury, whatever the investment rules laid down in a string of White Papers dictated.\(^6\) Clever economic rules about project appraisal, test discount rates, and required rates of return were often pushed aside when the overall budget constraint bit. State ownership now meant short-term investment horizons in a context within which longer-term planning was needed.

By the early 1980s, British Telecom (BT) provided a test case. BT wanted to invest in what were called System X exchanges. These passed the Treasury investment rules. But the new public-sector borrowing requirement (PSBR) constraints meant there was no money. Thus arose a central problem—perhaps the central problem which continues to motivate much infrastructure investment now—of how to invest without apparently increasing public borrowing. The borrowing constraint drove investment decisions, not the project’s economic value.

The answer in the BT case was first to propose a hybrid public–private bond (called the Busby bond). This failed the PSBR accounting rules test, in the sense that it could not be excluded, and so a second option was tried—privatization. In time all the main infrastructure utilities would be privatized—telecoms, airports, air traffic control, gas, electricity, water, and rail—and a gradual but massive transfer of responsibility to the private sector took place.

Privatization solved the public accounting problem in most cases (though rail partially came back on to the public accounts). But it did not solve the pricing, finance, and investment problems—as the nationalized model had also eventually failed to do. These required a new framework of regulation and customer billing, and with this the development of solutions to the contracting and time-inconsistency problems. We return to these below.

Not everything could, however, be easily privatized. Some areas remained stubbornly in the public sector, such as much of London’s transport network, roads, and flood defence. In other cases, new forms of public procurement emerged. The state would still provide the services—and pay for them—but the infrastructure projects such as schools and hospitals would be delivered by the private sector. The trick in these cases was also to try to find a way of keeping these projects out of the public accounts as had been achieved in the privatization cases. From the 1990s onwards, first the Conservative and then the Labour governments resorted to a new device—the PFI. To meet the accounting rules, the projects needed defined revenue streams. The borrowing and project finance would come from the private sector, and be paid for by charges for the services levied upon the future beneficiaries. These often turned out to be schools, hospitals, and local authorities—and hence proxies for taxpayers.

In both the privatized and the PFI cases, an additional argument was that current customers should not pay for future benefits. Rather the borrowing ought to be paid back by future customers—who would be the beneficiaries, whether as taxpayers or customers, and should therefore finance the projects. But while such considerations had an obvious rationale, each investment tended to be treated separately. The result was that no one questioned whether customers (and the public sector where it was the customer) could actually pay all these bills simultaneously. The investors relied on

these contracts to get back their fixed and sunk costs, but doubts would always remain about whether future politicians would actually make future customers pay. The cost of capital reflected these doubts, and it turned out to be an expensive exercise of shifting political risk on to the private sector—as we shall see.

Having parcelled out the infrastructure to the privatized companies and to private contractors under the PFIs, investment in infrastructure was now determined by a combination of new sectoral regulators (Ofwat, Ofgem, ORR, Ofcom, and the CAA) acting in concert with government departments, plus the Treasury in the case of PFI contracts. In a number of cases significant investment was delivered in this new largely private framework, notably in the water industry. But by the middle of the last decade, doubts began to emerge as to whether it was sufficient—whether the new arrangements would meet Britain’s needs.

As noted in the introduction, in part this was a function of new circumstances. In each of the main core infrastructures, for a variety of often distinct reasons, the investment requirement rose. Then there were wholly new infrastructures, such as broadband. Helm (2009) added up the total investment commitments by government to be delivered within the decade—by 2020—in the four main sectors—electricity, water, transport, and telecommunications. The result was around £500 billion, which represented, at £50 billion per annum, a doubling of the then current investment levels. The dominant part of this was in energy, to meet the decarbonization targets and the replacement of ageing power stations, but there remained a large investment programme in the water industry to meet the European directives, major road upgrades, new airport capacity, and telecoms and broadband rollouts.

These numbers were crude aggregations, and they left much out, particularly in respect of replacement capital expenditure. What was, however, unambiguous was that a step increase in investment appeared necessary, whatever the exact requirement. How then could this be delivered?

The incoming Coalition government in 2010 recognized the problem, and set about facilitating this additional investment, in a context that was matched only by that after the Second World War for the constraints on government resources. There are three planks to the development of the new infrastructure policies—planning, projects, and finance—in response to the perceived obstacles to delivering infrastructure: the barriers that planning created; the failure to prioritize projects; and the unwillingness of the private sector to come forward with the money. These proposed ‘solutions’ were brought together in the 2010 ‘National Infrastructure Plan’ (NIP) and the 2011 NIP (Infrastructure UK, 2010, 2011), and a series of subsequent updates (HM Treasury, 2012a,b, 2013).

(i) Planning

It has been a perennial and widely held view that the planning system stands in the way of infrastructure investment—typically without much by way of evidence. The argument boils down to two separate components: that it takes too long and costs too much to get planning permission; and that it results in decisions that go against particular apparently worthwhile projects. The former is about process; the latter is about outcomes.
It is true that planning takes time. In some notorious cases the length of the planning process has been extraordinary. Examples include the Layfield Inquiry into the Sizewell B nuclear power station in the mid-1980s (Department of Energy and Layfield, 1987), and the Heathrow Terminal 5 in the 1990s (Department for Transport, Local Government and the Regions and Vandermeer, 2001). But for all the notorious cases, there are a very large number that go through without undue delay. Indeed, the characteristics of the delayed cases are that they are contentious. Time and cost are related, but in most cases the costs of planning are small compared to the scale of the projects themselves. Time limiting has always been an option for reform. Cost could be further reduced by limiting what is taken into account, and the reforms to planning duly cut the guidance down considerably. But on these considerations, while there is scope for incremental reform, it is very hard to conclude that it is the planning process that explains the investment gap.

The real objection to planning is in relation to those projects which do not get the go-ahead. The reasons may be that they are simply not good projects, but there are other factors at play. Central to the decision are the interests that are taken into account and, in particular, whether local considerations dominate national interests.

The new planning regime built on the reforms of the previous Labour government, and in particular the creation of the Infrastructure Planning Commission (IPC) by the Planning Act (2008). The idea was that this body would identify nationally important projects and give them priority over local objections. The incoming government abolished this, but kept the central idea that there would be sectoral national plans. There would be a ‘presumption in favour of sustainable growth’ in the new National Planning Policy Framework with National Policy Statements (NPSs) for all the main sectors. In due course these have been published. The local losers would be overruled. The purpose of these statements could have been to set out the systems requirements for each sector, but this more coherent approach was stymied by the fact that these were planning statements, not policies. They summarized the existing policies. Missing were the underlying policies for the sectors. That had to wait for White Papers and other initiatives, or build on what already existed. In other words, it was for the government to lay down its national policies, and planning would always be set within this policy context. The planning reforms therefore created a mechanism through which policies could be translated into development. The first NIP was supposed to provide this, but it was general, rather than specific, and it remained—and in many cases still remains—to set the policies.

(ii) Priorities

While the first NIP was concerned with the framework for infrastructure projects, the second (Infrastructure UK, 2011) focused on specific projects. Five hundred are listed in Annex C, of which 40 were deemed to be ‘of national significance and critical to growth’ (p. 9). Infrastructure policy had become a list of ‘things to do’. In part this

---

7 Department for Communities and Local Government (2012).
reflected a desire to promote infrastructure spending as a way to give the economy a Keynesian boost, as a short-term stimulus, and as a result the projects selected tended to be the ones which had the potential to get going quickly and create jobs. But it also reflected the limited life of governments, and hence the need to deliver within the electoral timetable.

The result was an eclectic list of projects, ranging from the widening of the A14 major road, through to the interceptor sewer for London. No doubt many had merit, but gone was any serious attempt to provide a system-wide view of what needed to be done. The Treasury would now work on the delivery of each of the priority cases, with the hope, no doubt, that the result would not only be more ‘growth’ but also some sort of coherent overall infrastructure. The link between the systems and the projects was, however, no longer a focus of attention. That was the most important shift between the first and second NIP.

It mattered what was—and what was not—on the list. Pride of place went to prestige projects such as HS2, but politicians lobbied for their own pet schemes. A reading of Hansard on the debate on infrastructure in the House of Commons on 12 February 2013 is illustrative (Hansard, 2013). MP after MP focused on the projects in their own constituencies. Other important projects were left out—such as the airport capacity in the south-east. The list had become all too predictably determined by politics and the political process, and it was at this point that the government’s infrastructure reforms lost their radical edge. The greater objective of having a ‘national infrastructure plan’ had failed.

(iii) Finance

The third component of the policy concentrated on finance. The argument was that somehow investors were reluctant to come forward to finance projects with positive net present values. This must, the argument ran, be because markets had seized up, with banks and pension funds reluctant to invest. To the puzzle as to why rational investors would not take up apparently profitable opportunities, was added the peculiar feature of the times—extraordinarily low nominal (and negative real) interest rates—which should in theory have encouraged an investment boom.

The government came up with a Pensions Infrastructure Platform (PIP) to bring pension fund monies into infrastructure investment (HM Treasury, 2011b). Quite what the market failure was to which the PIP was supposed to be the answer remains opaque. The rather contradictory argument presented was that pension funds were, on the one hand, reluctant to pay intermediaries (such as infrastructure funds) to manage the assets and, on the other hand, too small individually to manage their investments themselves.

The government skated over another possible explanation for the funding problems: that once the political and regulatory risk had been taken into account—and particularly time inconsistency—the projects on the list might not be net present value positive at all. The reasons might be various, but they included two which remain unanswered: that the government has not come up with a way of being a credible contractor; and that the revenues promised to be paid by future customers and taxpayers may not be forthcoming. Promising that the fixed and sunk costs would be honoured required more than fine words. It required a bankable mechanism.
In the first NIP in 2010 (Infrastructure UK, 2010) the government contemplated a number of alternative ways of credible contacting. By the 2011 NIP, as the investors sat on their hands, demanding greater assurance in respect of the political and regulatory risks, the government conceded that it might have to give government guarantees.

The surprising feature of these guarantees was that they were directed at the one bit of risk the private sector could (and should) manage—the construction risk—and not the time-inconsistency problem. The concept built on the eventual outcome of the problems faced by HS1. In effect, the government underwrote the construction phase and then sold out its share on completion.

A final twist in the reforms on the finance side was PFI. The incoming government promised to abolish PFI, claiming it represented poor value for money—which, indeed, it did, as numerous studies and inquiries had shown. So the government abolished it—and then reinvented it in a remarkably similar way, but with the added twist that now the government would take stakes in the projects and hence share in the returns.

The logic was tortuous. The reason PFI was invented was to get major infrastructure projects out of government accounts, and to tie the projects to revenue streams—even if these were ultimately from the taxpayers. The arcane public-sector accounting rules drove the model. However unconvincing the argument, somehow they were then removed from the public expenditure numbers.

But the problem of public accounts was not solved by abolishing PFI, and in the bizarre world of these accounts, it turned out that PFI had to be reinvented. Sharing in the proceeds changed very little in terms of who paid, although by having the government with ‘skin in the game’ the chances of opportunist political interference through ex post interventions was somewhat reduced for the other investors. This risk was very real: the current Coalition government had first engaged in a direct attempt to change the terms of existing contracts, demonstrating graphically to investors the time-inconsistency problem of contracting with government.

IV. Reforms needed

It is not surprising that the latest enthusiasms for reforms of infrastructure fall a long way short of a coherent infrastructure policy. The three key components of a more coherent approach are: deciding what infrastructure systems are required; sorting out the revenues which can underpin the investments; and providing a credible contracting mechanism to overcome the government’s time-inconsistency incentives.

(i) Deciding what infrastructure is required

As argued above, infrastructure comes in networks and systems, not discrete projects. The systems come before the projects, not the other way around, as in the latest NIP.

---

9 See for a sample, House of Lords Select Committee on Economic Affairs (2010), NAO (2011), and House of Commons Treasury Committee (2011).

10 See HM Treasury (2011c).
updates (HM Treasury, 2012b, 2013). The National Planning Policy Framework is a start in the right direction, but the NPSs are largely summaries of existing policies. What is missing is coherent sectoral policies within which the planning statements can sit. The 2011 NIP (Infrastructure UK, 2011, section 3) recognizes this, but when it comes to setting these out, it is about ‘vision’ and ‘ambitions’ rather than decisions. It is not the planning process that is the problem, but the absence of decisions about the sectoral systems.

The position varies from sector to sector. In water, the EU Water Framework Directive and a series of related European directives set the scene for the determination of the required investment. This is mediated by the Department for Environment, Food & Rural Affairs (Defra) and the Environment Agency (EA), and the regulator, Ofwat, is then charged with ensuring that it is carried out in a cost-efficient way. Though there remain serious problems—notably with over abstraction and the role of abstraction rights—water is the sector which comes closest to having a coherent overarching policy framework.

In aviation, the absence of a policy is widely acknowledged. It is for government to decide what airport capacity is needed and where it should be developed. In particular, where demand presses tightly against supply in the south-east, there are clear choices to be made—whether to add runways to Heathrow, Gatwick, or Stansted, or to build a new airport in the Thames estuary. The trouble is that they are not made, and the result is the congestion. It is also for governments to decide on the rail and road links to airports. Again the decisions are not made—not least because it is hard to design a high-speed rail system and the future development of the strategic road networks unless one knows where the airports and runways are going to be.

In rail and roads, governments are forced to choose systems. In rail, the mechanism is to decide what spend the government will subsidize through the train-operating companies and through to Network Rail. It is for government to decide whether to electrify key lines, and to decide about major interconnections. Government decides about Crossrail and whether it will be supported by new north–south London routes. With government guarantees, the rail case is one—like water—where government is, indeed, forced to decide. In roads, it is direct government finance, so again it decides. The decision on HS2 has in principle been made, but uncertainty remains, not least because there has been a history of policy statements being abandoned. The classic example is the 2000 Transport White Paper (DfT, 2000) in which the objective of switching 50 per cent of traffic from road to rail was announced.

In energy, the choice of system is fundamentally a contested one. If the decision is to move towards decentralized, intermittent, and remote sources of low-density power, then the required network is the mirror opposite of what currently exists—designed for large-scale power stations with base-load capacity. The EU Renewables Directive (Directive 2009/28/EC) dictates a path to 2020. However, there is little clarity about what happens after 2020, and government has a series of decisions in its hands which will determine what sort of system is needed. For example, if a series of nuclear power stations are built, providing large-scale base-load low-carbon energy, then they will need one sort of transmission system. If, on the other hand, they are not built, then obviously a whole chunk of transmission will need to be different. Here is a classic system choice to be made.

In communications, it is understandable that rapid technical change has made it hard to work out what a coherent system might look like. However, some things are already
apparent. Broadband is now a new and essential infrastructure—needed for business, and for any individual to participate in society. Furthermore, as in many such systems, some customers are more profitable than others, but all benefit if everyone else is connected. There is then a high-level decision to be made—the level and coverage of the broadband system. An aspiration to have everyone connected to high-speed broadband is not a decision. A decision is about the system itself.

The result across these major infrastructure sectors is at best a mixed one. There is a framework for water and railways, and roads are in the public sector anyway. There is little clarity around energy, airports, or broadband. Unfortunately these are key systems and represent a considerable bulk of the investment requirement. The first necessary step to sorting out Britain’s infrastructure is to decide about airports, about nuclear power and the energy mix, and about broadband systems. Perpetual failures to do so reflect deep problems in the political process and the institutions for binding governments to longer-term commitments, and ensuring that revenue streams are sustained.

(ii) Sorting out the revenue streams

Much of the policy agenda has been about indentifying sources of finance. Though there may be problems associated with financial markets following the credit crunch and ensuing banking crisis, there is, in fact, little evidence to suggest that it is a quantitative constraint on investment in infrastructure. The provision of finance is, on the contrary, about price not quantity, and that price—the cost of capital—is a reflection of the fact that the revenue streams are often at best uncertain.

Why might the revenue streams from new roads, energy networks, sewers, and high-speed rail networks be uncertain? There are three questions investors might ask. Can the customers (or taxpayers) actually pay? Even if they can pay, will they vote for politicians who will force them to pay? And, will the asset be stranded by technological progress?

Let’s start with the potentially binding constraint: the ability to pay. Let’s also consider a specific example: energy. By 2015, it has been suggested that perhaps 25 per cent of all households will spend over 10 per cent of their disposable income on household energy supplies. Current bills include around 10 per cent levies and subsidies for fuel poverty measures, renewable, and related costs for the chosen path towards the renewables energy target. This is the beginning of a ramp up in bills for what are some of the most expensive ways to make marginal reductions in carbon emissions. The result is that energy bills are now recorded in opinion surveys as among the most important political concerns. In consequence, there is a levy control mechanism which limits the total subsidies in Britain (HM Treasury, 2010), and most European countries have

---

11 A recent example is provided by Alistair Buchanan, Chief Executive of Ofgem who blames the lack of investment in energy on the credit crunch, rather than the failures of Ofgem to set the appropriate incentives (Buchanan, 2013). In fact, the credit crunch reduced demand for energy, and therefore improved security supply.


13 See Helm (2012, ch. 4).
taken measures to reduce subsidies. In the case of Spain this has even had a retrospective element.

Investors in the transmission assets, in nuclear power stations, and in offshore wind in particular, have taken note. In the last case, high discount rates enable the investors to get their capital back quickly, and in a number of cases to exit by selling on the completed assets, like wind farms. This rational response—a high cost of capital—has consequences: the high cost of capital raises the costs to consumers, who then are less able to pay and use their political voice to oppose the subsidies. This raises political risk further, raising the cost of capital.

Customers do not consider energy bills in isolation. Their concern is the household budget as a whole. They face rising water bills, higher transport costs, and the costs of new services via broadband as well. To this aggregate utilities bill is added local taxes, mortgage costs, and the food basket. Currently, investors will note that mortgage costs are extremely low by historic standards. Thus they must factor into the ability to pay the consequences of mortgage rates rising to more normal levels—based on interest rates in excess of the rate of inflation. If, say, 65 per cent of households are owner-occupied and currently mortgage payments take up around 20 per cent of household income, then a doubling of the nominal cost of mortgages, added to the rising bills for energy and the other utilities, results in the obvious but largely ignored conclusion that the aggregate implies a significant fall in living standards and is probably unsustainable—especially politically. If it is unsustainable, it will not be sustained. For an investor, the conclusion is obvious—use a high cost of capital hurdle so that the focus on returns is short term. Better still, get the government to be a co-investor and pressurize regulators to ensure that customers pay for assets in the course of construction, as they did under the nationalized system. For many private investors, the result may simply be that they do not invest.

The ability to pay is a necessary condition for private-sector finance. But even if customers are willing to pay, they might not vote for politicians who force them to pay via the levies and regulatory system. Politicians react in the political market to voter preferences. Current voters have a strong incentive to behave in a time-inconsistent way, and demand marginal cost pricing. Politicians respond, and they have the ability to legislate and hence change property rights. Investors seek legal protection, but governments have multiple ways of impacting on investors ex post. For example, while constitutional rights (especially European ones) limit the ability of a government retrospectively to change contracts and levies, they do not protect against other actions. In the British context, ex post windfall taxes have been resorted to. In Germany, a nuclear fuel rod tax was introduced. There are a host of other ways governments can influence costs, including business taxes. Again, given these political risks, investors look to a quick pay-back of the sunk costs, and hence use a high discount rate.

A final reason for using high discount rates is that there may be technological change. The longer the life of the asset, the greater is the exposure. This can come quickly, as in the case of mobile telephony and the internet, or more gradually as existing techniques are improved. Building a nuclear power station with a life of, say, 60 years exposes investors to the possibility of a host of new generation technologies. Indeed, in the nuclear

---

14 On the British 1997 windfall tax on the utilities, see Helm (2010a).
case, the coming of shale gas in the US has already massively impacted on the economic case for new build. Unsurprisingly again investors seek government guarantees.

These considerations point to a gap between the private and public investment horizons. There are good reasons for thinking that the social discount rate should be quite low. The Treasury Green Book suggests a real rate of 3.5 per cent.\footnote{HM Treasury (2003) and HM Treasury (2011a).}

Once decisions about systems have been made, infrastructure policy should concentrate on this issue—bridging the gap between the private and social costs of capital—rather than the misguided attempt to address an assumed quantitative financial constraint. This requires a credible contract that is proofed against time-inconsistent behaviour.

(iii) Providing a credible contract

Investors know that government (and voters as taxpayers) have a powerful incentive to behave in a time-inconsistent way, and they know, too, that governments have an incentive to commit to future aspirations and targets the consequences of which they, with their limited time horizon, will not be around to face. Targets for 2050 (the EU Energy Roadmap (European Commission, 2011)), the 2020 renewables target (European Parliament, 2009), and the HS2 investment horizon to 2032 (DfT, 2013) are particular examples of a general phenomenon. Even the monetary union and the eurozone were constructed on this basis. Government knows that investors know this. So if government wants investment to take place, it has to find a way of convincing (or even fooling) investors that it will not resort to time-inconsistent \textit{ex post} expropriation.

It is important to realize just how difficult it is to make credible political commitments. For much of the last century, such a contract could not be found. Nationalization was the consequence: since governments could not commit to private investors in a credible way, they had to do the investing themselves. To put it another way, they could not find a contract which could close the gap between the social and private discount rates.

In principle, there is a variety of ways in which the gap could be closed. First, the government could sell monopolies to the private investors. The ownership of the monopoly would enable the private sector to hold the government to ransom in the future, and hence by owning valuable assets free from competition, they could force the customers to pay. This is, indeed, how the early railways in Britain and the electricity networks in the US got built.\footnote{See Yergin (2011, ch. 17).}

The problem with the sale of monopolies is that even this is an incomplete contract. It typically does not exhaust the element of governmental control. The government can regulate prices, and typically does so—either through direct regulation or through competition law (which provides that dominant positions should not be exploited). It can tax, and it can make life difficult for an incumbent in a host of other ways, given the numerous interactions between private companies and governments.

A second approach is to offer a whole-of-life contract for a specific asset. This is the PFI route. The private sector is asked to build an asset, and is then guaranteed a revenue stream subject to maintaining the assets and delivering the outputs. The protection
for the investor is that the contract specifies the revenue stream and that there will be no
\textit{ex post} changes to that contract. Again, however, it is not so simple. Outputs are rarely
completely specified over, perhaps, decades. Things change. Renegotiation is inevitable. The private investor in PFIs typically has multiple relationships with the client—government. So when government realizes the political consequences of the costs of the revenue streams the contracts impose, it launches inquiries and demands rebates.\footnote{HM Treasury (2011c).} The threat is real: unless the investors ‘compromise’, further deals are in jeopardy. And so on.

Unsurprisingly, therefore, the evidence from PFIs is that the gap between the public and private discount rates is not closed, and the ability and willingness to pay on the part of future taxpayers is rarely properly analysed in advance. Investors, therefore, demand a high cost of capital, and the \textit{ex post} inquiries and interventions merely act to reinforce this.

A third option is to use the RAB model.\footnote{See Helm (2010b).} Here the assets are transferred at privatization to a corporate entity, typically a profit-seeking company, but sometimes a not-for-dividend company or some form of trust. The private entity is granted a licence and is required to fulfil the functions as defined in that licence (or related contractual form), and in return an ‘independent’ regulator is required to regulate prices and outputs to ensure that the entity can finance its functions. The regulator is a creature of statute, at arm’s length from government, which gives the contract some credibility, though even here it is relative.\footnote{As witnessed, for example, when the shareholders of Railtrack accepted the demise of the company.}

In order to ensure the functions are financed, the regulator needs to know what the investors’ stake in the company is, and this falls into two broad categories. There is the requirement to recover the operating costs, and the requirement to finance and recover the capital expenditure costs. The former is straightforward: these are recurring and continuous costs. The latter falls into two further subdivisions: the finance for the capital expenditure in the course of construction, and the recovery of the fixed and sunk costs once the project is competed.

Here is where the RAB comes in. At the price-setting periodic reviews, the regulator determines whether the capital expenditure in the previous period has been efficiently conducted. It then transfers the (efficient) number into an asset base—the RAB. This asset base is the aggregate of the efficient capital expenditure plus an initial value for the purchase of the existing assets. This is what is exposed to time inconsistency: it is fixed and sunk, and it is inert, in the sense that there is nothing that managers can do to change it. If the regulator guarantees that customers will pay a return on this, then it is customers, not investors, who bear the equity risk, and it can be debt financed. If it is further protected through indexation from inflation, it is akin to an index-linked government bond.

This is precisely what has emerged almost by accident in most of the British regulated utilities. The RABs solve the time-inconsistency problem, and therefore close the gap between the government and private costs of capital for these sorts of assets. Utility bonds have the RABs as collateral (since the companies own the assets) and they have
statutory protection via the regulator. The result is that the RABs are overwhelmingly debt financed at rates close to the government’s cost of debt. Indeed, circumstances might even be envisaged in which these RAB assets are less risky than government bonds, since they are asset backed. Unsurprisingly, this is what the overseas investors in British utilities—which the government is so keen to court—have found so attractive: existing RABs, rather than greenfield new investments, and investments in infrastructure which then roll into RABs.

In the first NIP the government considered the RAB model and undertook to study it further (HM Treasury, 2010). However it raised a number of objections to generalizing the RAB model. The Treasury states that: ‘as the RAB model passes the risk of paying for sunk costs in respect of infrastructure investment on to the consumer, consideration must be given to the affordability implications of this approach’ (pp. 15–16).

It then goes on to make the surprising claim that: ‘Applying a RAB model to assets which are delivered within a competitive market is likely to result in the removal of the advantages of competitive pressures for greater efficiency, optimal operation and innovation.’

Both of these objections display an extraordinary lack of understanding of the issues. Sunk costs are sunk: they do not go away because government might not want customers to pay. If they do not pay, either investors will be expropriated or taxpayers will pay. But even more surprising is the lack of understanding of what a competitive market is, and in particular confusing markets where there are competitive elements, with markets where private incentives are sufficient to deliver optimal investments in infrastructure. It is precisely because these incentives are almost always insufficient, as set out in section II, that intervention is needed. The question is whether the RAB approach is better than the alternatives. On this the Treasury is silent.20 It appears unable—or unwilling—to recognize the time-inconsistency problem, let alone provide a credible contractual solution.

Notwithstanding the flaws in the NIP analysis noted above, the government reinvented the PFIs, and turned towards direct state guarantees as a route forward—in other words, edging back towards the twentieth century model of the government itself using its position to underwrite investment costs, to close the gap between the public and private costs of capital. The government did, however, continue to flirt with the RAB model, notably in the Prime Minister’s speech on infrastructure in February 201221 where he suggested that the water industry’s RAB-based approach might be extended to roads. Ironically, his intervention came at just the point when the water regulator was questioning the RAB concept, and seeking to undermine it.22 Such examples not only reflect the lack of coherence and consistency in British infrastructure policy, but further act to undermine private-sector confidence, and hence drive up the cost of capital.

20 The NIP goes on to say that ‘The government will conduct an internal review, supported by external experts, to consider extending the use of the regulatory asset base model’. As one of the ‘external experts’ the author is unaware of any outcome, report, or assessment that may have resulted. In the NIP 2011, the Treasury merely notes that it had considered the extension of the RAB model to strategic roads and flood defence, but cannot make up its mind about a revenue stream. This indicates a problem with deciding about road tolls and flood defence charging—not a problem with the RAB model.

21 Cameron (2012).

22 The water position was further confused by a statement from the Chairman of Ofwat in March 2013, in which he challenged the capital structure of the industry, proposed a possible form of profit sharing, and flirted with ex post clawbacks and voluntary tax contributions. See Cox (2013).
V. Return of the state

The flaws in the government’s approach to infrastructure policy have gradually become apparent. The cost of capital gap has not been closed—indeed it has, if anything, widened. A number of current investment projects reveal this. Nuclear is a good example. The asset is long lived, and the marginal cost once it is operating is very low, opening up a big gap between the marginal and average costs. The lead investor for the first new nuclear power station is Electricité de France (EDF), which is 85 per cent owned by the French state. It considers that a guaranteed price for its output is required which should meet its cost of capital. It is widely reported in the press that EDF thinks that, with the absence of government guarantees, this lies between 8 and 12 per cent. In contrast, the government’s current cost of debt is negative in real terms. This gulf in the cost of capital swamps everything else—even the possibility that the costs of the project might be higher and it may take longer to build. Indeed, cost escalation and delays are much worse at a higher cost of capital.

The gulf could be closed in a number of ways. The contract could be more generous: there is a price of electricity which would make the project economic. But customers might not pay it, and the higher it is, the less likely it is to be sustainable. Alternatively the government could underwrite the risk. As with HS1 and HS2, the idea is that it is the construction risk which needs support—precisely the risk that ought to be in the private sector. A third option is to turn nuclear into a utility, with a RAB to protect the sunk costs. The neat feature of this approach is that it is the RAB costs that customers would guarantee to pay, rather than a fixed price of electricity, and only when the project is finished. What would go into the RAB would be the efficient costs of the building the plant, so managers would have the equity risk of delivering to time and to budget. The model could be extended to wind farms and other renewable energy technologies, where there are sunk costs and political and regulatory risks, and hence investors are exposed to time inconsistency.

But having rejected this RAB approach, the government is now faced with the consequences. Since it cannot allow the infrastructure not to be built, and it has rejected the routes to closing the gulf between the private and public costs of capital, it will now either have to take on the political and regulatory risks itself, or not much of the investment will take place. By default, government is back.

So far there have been three steps in this direction: subsidizing the private sector; giving guarantees; and state ownership. Subsidies have now been advanced to water customers in the south-west on the grounds that they are no longer able (or politically willing) to pay (HM Treasury, 2011b); direct funding of broadband rollouts in rural areas has been provided; and direct ownership proposed for HS2 (DfT, 2013). Indirect subsidies via levies on utilities have also become widespread.

Guarantees have been advanced to Network Rail, and some £45 billion has been indicated for guarantees to some on the government’s priority project list (HM Treasury, 2012a,b). Direct ownership applies in the case of London Underground and Crossrail. This is a process that is likely to accelerate. It is possible that the government may even take ‘stakes’ in core infrastructure as it has already proposed in its revised PFI approach.

The return of the state to infrastructure is being reinforced by a number of institutional mechanisms. There is already a Green Bank, and a number of proposals from a broader based National Infrastructure Bank. The Bank of England has bought significant quantities of utility bonds as part of its quantitative easing policy.
It remains to be seen how far this trend will go. The current constraints on public borrowing limit interventions, but in part this is a function of an accounting failure. Deficits are measured in cash terms. The government has no balance sheet, and a case can be made that investment in new assets should be set against the borrowing liabilities, and if capital maintenance is deducted to maintain the value of assets intact, it would provide the basis for a sustainable approach to both public finances and infrastructure investment (Helm, 2011).

VI. Conclusions

The recognition of the importance of infrastructure by all main political parties has brought with it the problem of delivery. Britain has tried most models over the last century: unregulated private-sector activities, municipal ownership, state ownership, monopolies, competition, privatization, utility regulation, PFIs and variants such as PPP, direct subsidies, and state guarantees.

In the developments of infrastructure policy in recent years there has been a focus on encouraging private finance, notably from sovereign wealth funds and pension funds. This emphasis has put the cart before the horse: the problem is not one of quantitative constraints on the flow of funding, but rather the nature of the risks of infrastructure investment and the failure to close the gap between the public and private costs of capital.

This failure has arisen because governments have focused on the immediate problem of getting investment off the public expenditure accounts, rather than on reducing the political and regulatory risk, and this explains part of the gap in the cost of capital. That risk relates primarily to the time-inconsistency problem. At the current private costs of capital, outside the RAB-based utilities, investment is falling short of aspiration because customers may be unable to pay, or unwilling to pay. Governments have so far failed to come up with a credible way of convincing investors that they can make credible commitments outside these core regulated utilities.

The problem is not complicated or insurmountable. The RAB-based model has shown its worth—protecting the fixed and sunk costs while placing the project construction and operation risks with those best able to manage them. The government argues that this model only works in the customer-only financed cases. This is wrong: indeed, the RAB-based model is even more relevant where the government is involved directly on behalf of taxpayers. It is in these cases that the incentives to behave time inconsistently are greatest, and where contract credibility may be weakest.

The result of this failure to embrace the RAB model is a gradual reversion to the twentieth century model—state guarantees and forms of state ownership. The longer-term implication is that the great privatization experiment may have run its course. The state is back—with all the inefficiencies that this may bring.

References


— (2012), The Carbon Crunch: How We Are Getting Climate Change Wrong and How to Fix It, New Haven, CT, Yale University Press.


