The sustainable borders of the state

Dieter Helm*

Abstract Economic crises have historically set the stage for the reconfiguration of the economic borders of the state. Current concerns focus on debt and associated intergenerational issues, such as pensions, infrastructure, and the environment. The sustainable state is designed to put in place a framework for integrating these intergenerational issues, alongside the traditional focus on macroeconomic policy, static market failures, and transfers. The building blocks include the accounting framework and, in particular, a national balance sheet setting out assets and liabilities, and providing for capital maintenance over time. To support these assets and to provide for sufficient investment, the sustainable state requires an appropriate level of savings, and institutions to translate these savings into investments. Two components are: the deployment of regulatory asset bases to underpin long-term sunk and fixed costs, given the time inconsistency problem in government’s commitments; and the development of financial intermediaries such as an infrastructure bank.

Key words: government performance, macroeconomic aspects of public finance, state and local taxation, subsidies, revenue

JEL classification: H11, H170, E60

I. Introduction

Each generation inherits the economic borders of the state from its predecessor. Major changes—such as those initiated by the Attlee, Thatcher, and now the Coalition governments—involves creating both winners and losers. Losers tend to shout louder than winners: hence the existing economic borders of the state are protected by lobbyists and vested interests and are very difficult to shift. It is therefore unsurprising that the major shifts in the economic borders of the state have followed wars and economic crises. Now we have a major economic crisis, and hence one of those rare historical moments when the design of the state itself is, across the developed world, a variable, and the debate is about more than how to make incremental improvements to the largely given economic borders.

In the second decade of the twenty-first century, 5 years into this economic crisis, the key challenge is not only to deal with the immediate issues of debt and the deficits, but also the wider issues, from pensions and care for ageing populations, through to the environment. Not only is it widely agreed (at least politically) that current debt levels and deficits are unsustainable, but the provisions for future pensions and health, and the protection of the
atmosphere and biodiversity have moved from being specialist interests into the mainstream of political and economic debate. They share one key feature: they are all long term and intergenerational in character. The economic borders of the state—and the supporting ways in which social provision, the environmental interests of future generations, and the infrastructures linking the generations are incorporated—are currently not designed with this intergenerational perspective to the fore. Rather the economic borders of the state have evolved with a concentration on managing the business cycle and inflation, social transfers and welfare provision, and microeconomic market failures focused on competition and labour supply (Helm, 1986, 1989).

The structure of this paper is as follows. Section II addresses the twentieth century legacy, explaining why we have ended up with the state in its current form and with its current scope. What were the questions to which the twentieth century state was supposed to be an answer? Section III turns to the new objectives, and in particular the role of sustainability and intergenerational considerations. What are the questions to which the twenty-first century state is supposed to be an answer? In section IV, the redesign of the state is considered. What would be a better answer for the twenty-first century? Section V concludes.

II. The twentieth century legacy

The historical growth of the economic functions of the state has not been smooth: rather these have been characterized by long periods of stability, punctuated by substantial expansions. Setting the scene for the twentieth century, the Victorian state was a product (and a cause) of trade, empire, and the industrial revolution, with a gradual expansion of municipal and central government involvement in infrastructure, and early welfare provision. It was never a ‘night-watchman’ state, conjured up in the theories of laissez-faire. Capitalism could not develop and thrive in the absence of a fairly forceful centralized state, which was not only engaged in enforcing contracts and property rights, maintaining the gold standard, and policing the high seas, but also interfered in a host of industrial and infrastructure matters. It was never the state or the market, but the state facilitating the market. What Hayek called ‘the constitution of liberty’ was not laissez-faire (Hayek, 1960).

But these interventions were limited, and mainly focused on the framework within which markets operate, not replacing the markets themselves, and there was a widely held consensus that they should be limited. Free trade, free labour markets, and market prices were matters where government generally should keep out. The business cycle was regarded as inevitable, although governments gradually took greater responsibility for the relief of the poor and the regulation of factory labour, and local authorities stepped in with sewers and public health measures. By modern standards these interventions were very primitive.

By the end of the nineteenth century, the state in economic terms remained very small.¹ The First World War changed this, in response to the unprecedented requirement for total industrial mobilization—a task beyond the private sector. General government expenditure doubled to around 25 per cent of gross domestic product (GDP). War brought taxation from

¹ In the US, general government expenditure was less than 10 per cent, and it was about 10 per cent in the UK. See Table 1.1 in Tanzi and Schuknecht (2000). See also Tanzi (2011).
the sidelines to centre stage, and unintentionally caused major social change: the classes were forced to mix, and the sheer scale of the carnage altered political and economic perspectives.

But for all the trauma of the First World War, the extraordinary consequence was that the immediate post-war period saw a strong desire to return to ‘normal’, by which was meant the Edwardian period. After the war and demobilization, the economy recovered, and in due course the Gold Standard was re-established. Labour relations soured, but the General Strike in 1926 did not make much difference. The conventional view of the state was still a nineteenth century one, albeit that Britain had piled up large debts, and had lost out significantly in its trading dominance—in part because trade did not fully recover, and in part because the US (and others) were now serious competitors. Britain had an expensive empire and lots of debt, and its industrial base could not sustain them. The belief in a minimal state and the reality differed: in fact, the share of government did not fall back despite the nostalgia for the past.

When the Great Crash hit the US in 1929, the belief in the minimal state, to the extent that it ever existed, was comprehensively undermined. It set the scene for an economic debate which has shaped economists’ views about the economic borders of the state ever since. In the 1930s, there were two new strands to the debates about the economic borders of the state: the role of the government in managing the business cycle (and the Great Depression was a massive business cycle); and the role of the state in improving economic efficiency. The former is now associated with Keynes and the Keynesians (versus the monetarists); the latter is associated with what was then the ‘new’ welfare economics, based upon the Pareto principle, and subsequently morphing into market failure analysis.

The debate about intervention in the macroeconomics of the business cycle and in the microeconomics of market failures was not translated into a step change in the role of the state until a further war intervened. The Second World War lifted the US and Europe out of depression, and this was largely at the direction of government. In the Second World War, the state regulated most aspects of the economy, and in 1945 there was no going back to a limited role for the state—at least in Europe. State direction of labour, industrial policy, and the command of economic surplus remained. Indeed, rationing persisted into the 1950s, and British government debt exceeded 250 per cent of GDP—much greater than in the current economic crisis.

New layers were added: nationalization placed the commanding heights of the economy in state hands; and the welfare state took over a large swathe of responsibilities for health and education that had previously been left to individuals and non-governmental organizations, and took the recommendations of the Beveridge Report on board (Beveridge, 1942). Farming was to remain largely planned (and subsidized) for the rest of the century, and town planning now became a state function. Full employment became a primary objective of government.

It took until the 1960s for the full implications of these new objectives to play out. The post-war boom meant that the full employment objective could be met (arguably largely independent of government policy). Across Europe, growth rates exceeded previous historical experience. Even in Britain, economic growth (real GDP per head) between 1950 and 1973 exceeded 2 per cent p.a. compared with the longer-run rate in the nineteenth and first half of the twentieth centuries (1820–1950) of around 1 per cent p.a. Catch-up was an important part of this leap forward.² The expansion of the state into industry in the 1940s

² See Eichengreen’s (2007) convincing explanation of the role of catch-up in Europe after 1945. Table 2.2 (p. 17) provides real GDP per head 1820–2000 for European countries.
and 1950s played out in the large investment programmes for energy and transport infrastructure. The debt overhang was dealt with in part through economic growth and in part through a gradual expropriation of savings (backed up by controls on the banks and their obligations to hold government debt, and exchange controls which prevented capital flight, and topped up with (very) high taxation).

In the 1960s, the Beveridge Report’s preoccupation with absolute poverty gave way to a more relativistic approach, and the state’s share in the economy was inflated by transfers from around 30 per cent to over 40 per cent, where it has remained ever since, despite serious attempts to reverse its growth in the 1980s. By the end of the 1970s, the state had become both a producer and a transfer state. This was true for Europe, too, (east and west) and, indeed, even for the US.³

The 1970s were a period of economic crisis. The OPEC oil shocks combined with a period of labour unrest to produce stagflation. The combination of social unrest, trade union militancy, and a slowdown in economic growth, with inflation peaking at over 20 per cent, resulted in an economic crisis which in turn precipitated attempts in the 1980s across developed countries to roll back the economic borders of the state. Competition, privatization, liberalization, and monetarism came to replace monopoly, state ownership, and Keynesian demand management, most notably in the policies loosely associated with Reagan and Thatcher, but subsequently followed across Europe. The collapse of the Soviet Union at the end of the 1980s provided a further impetus towards more market-orientated economic policies.

The most notable feature of this attempt to roll back the economic borders of the state is how little it actually achieved in terms of the share of government. It changed its functions but not its scale nor its overwhelming influence on the economy. While privatization transferred major industries into private ownership, the degree of state control remained significant—now through the growth of regulation and the regulatory state. Welfare spending remained largely immune to the more market-orientated approaches and, indeed, continued to grow both in its share of government expenditure and in its complexity.

Few doubted that this broad balance would be maintained: the ambition to roll back or radically alter the economic borders of the state remained only with relatively isolated groups on the left and right. What changed all this was the economic crisis, which manifested itself from 2007 onwards, but which actually started much earlier. It was very much a macroeconomic event, but with deep microeconomic roots, and the initial Keynesian responses led to a large increase in the stock of debt and deficits. It gradually became apparent that this was not a short-term dip but a longer-term shock, and both the debts and deficits were no longer sustainable. Quite quickly, the conventional wisdoms about the economic borders of the state and, in particular, the assumptions about the sustainable level of debt and consumption came under intense scrutiny.

The crisis has had its phases, but its origins lay back in the great boom of the 1980s and 1990s, and in the policy responses to its ending, notably as stock markets turned down around 2000. The macroeconomic policy response was then very Keynesian: both fiscal and monetary policy was relaxed. Real interest rates were pushed down towards zero in the US, the UK, and Europe. Fiscal expansion—tax cuts in the US and a major increase in public expenditure in the UK—accompanied the monetary stimuli.

³ See the paper by John Hills (2011, this issue) for a discussion of how the architecture of the welfare state has evolved in recent years.
The impacts were not immediately apparent, but by lowering the real cost of debt after 2000, both the level of borrowing and the switch from equity to debt were stimulated. The dash-for-debt was extraordinary on any historical basis. Savings ratios were well below sustainable levels and asset prices (especially house prices) rose strongly. Not only were the returns from savings low, but also by increasing leverage everyone could try to benefit from rising asset prices. Consumption was the key driver of growth in the economies of the US and Britain. The living standards were increased on the basis of borrowing against future income. Consumers were taking an increasingly large mortgage on the future, and in the corporate and banking sector the debt leverage played a similar role.

By 2006, real interest rates started to rise tentatively, exposing the extent of sub-prime lending, initially in the US mortgage market. An (inevitable) banking crisis followed, crystallized in the collapse of Lehman Brothers. To prevent economic collapse, governments stepped in to implicitly or explicitly take over much of the banking system and its liabilities, interest rates were further reduced, and government spending increased. The examples of monetary and fiscal stimuli after 2000 were repeated on a much larger scale.

As the crisis unfolded, government deficits increased across the developed world. Tax revenues fell, while unemployment and other transfer payments rose. To many mainstream Keynesians, this was the moment to really test their theory: government should accommodate, and indeed stimulate, to offset the contraction in the corporate and domestic sectors. Once real interests rates had been driven to zero, quantitative easing plus fiscal stimuli could (and should) fill the gap in aggregate demand. And if the economy failed to respond to the policy stimuli, then as long as there was excess capacity in the economy, the stimuli should be increased until it did respond. Keynesians continued to advocate tax cuts and oppose spending cuts.4

Once the initial credit crunch had passed, policy-makers increasingly ignored the Keynesian prescription, and instead focused on reducing deficits. The reasons were obvious. The problems with increasing government deficits were: first, that there had to be a willing lender if governments were to be willing borrowers; and second, the private sector had to respond to the stimuli. Neither turned out to be completely compliant. The bond markets drove up the cost of government borrowing in a number of countries (notably in the eurozone) and threatened to do the same to the US and Britain. For as long as the British government presented a credible plan for deficit reduction, lenders proved willing (even in the context of a possible future devaluation). In the US, a similar position pertained. But what the Keynesians failed to explain is what would actually happen if governments changed tack and deliberately cut taxes and increased expenditure. It is not hard to imagine a crisis: a sharp fall in the exchange rate, rising inflation, and a rise in interest rates—or default.

The Keynesian response is that fiscal stimuli would be self-financing: the private sector would respond, the ‘multiplier’ would eat into the spare capacity, and hence the stimuli would reduce the deficit.5 The problem for them was deficient aggregate demand. The response should be to increase it. This mechanistic view of the private-sector response reveals a profound difference of view: for the Keynesian starting point is the level of consumption in 2006—before the crisis hit. It is implicitly argued that this level of

4 See the following articles in this issue for a discussion of some aspects of these questions: Frankel (2011), Gale and Harris (2011), and Sapir (2011).

5 Evidence for the multiplier effect remains elusive, since the counterfactual—what would have happened in the absence of the stimuli—is unobservable. See Ramey (2011) and Taylor (2011) for surveys of the evidence.
consumption can and should be sustained. An alternative view is that the position in 2006 was unsustainable: that savings were too low, the core infrastructure was being consumed rather than replaced, North Sea resources were being depleted, and the corporate sector was focused on financial engineering rather than physical investment. The performance of the economy was flattered by the financial sector—whose earnings were based on leverage. In other words, the alternative view is that the level of consumption in 2006 was unsustainably high, and would have to fall. It was an illusion based on debt, and represented a transfer from the future to the present.

This debate matters not only from a macroeconomic policy perspective—it has profound implications for the economic borders of the state. It turns attention towards what level of government activity is sustainable, and the extent to which the underlying consumption level rests on a framework that does not prejudice consumption in the future. While Keynes could focus on the short run, and comment that ‘in the long run, we are all dead’, future generations might beg to differ.  

As part of the reappraisal that the economic crisis instigated, it is increasingly recognized that the future cannot be left to take care of itself, and that deficits, debts, depleted resources, and degraded infrastructure are a legacy which future generations might not thank us for. For while the economic crisis has focused policies on the aim of stimulating immediate economic growth, at a deeper level, intergenerational objectives have begun to intrude into the questions of the role and scope of the state, and its economic borders.

III. The new objectives and priorities—sustainability and international public goods

To summarize so far, the twentieth century witnessed the growth of the state in terms of managing the business cycle, in correcting market failures, and in transfers. The growth of transfers created tensions, notably between efficiency and equity. But notwithstanding these tensions and, indeed, the mixed results from trying to master the business cycle and enhance the efficiency of the economy, there has been no decline in the ambitions for the state, and in particular its responsibilities. Indeed, quite to the contrary, in the 1980s, environmental concerns added another dimension, as these shifted from the conventional debates about local biodiversity and habitat loss to encompass climate change. It became apparent that economic activity was changing the composition of the atmosphere, and that increases in greenhouse gases might change the climate, potentially by several degrees centigrade.

Since climate change was not something markets could be relied upon to correct themselves—it represented a public bad—governments were called upon to intervene. Now the debates about the economic borders of the state focused not just on efficiency and equity in the conventional senses, but also on efficiency and equity through time.

It is hard to underestimate the profound implications for the economic borders of the state that this intergenerational question implied. For intergenerational concerns involve not just the preferences of existing people (and therefore existing consumers and voters), but also future ones, too. While it might be a defensible heuristic to assume their preferences will be

---

6 Keynes's (1931) essay ‘Economic Possibilities for our Grandchildren’ illustrates this argument. More generally, it is possible to view much of the Keynesian policy prescriptions in a short-run perspective.
the same as ours—that human nature is, as David Hume had suggested, ‘always and everywhere the same’—the same cannot be said for their technologies.

The conventional approach to ‘the future’ in economics has been to assume that economic growth will continue, so that future people will be better off than we are. This, indeed, explains the low levels of savings witnessed in the UK and the US in recent years. The problem has been to make sure that we do not bequeath too much to them, if we are to be equitable through time. Hence their consumption is discounted by the assumed growth rate. Growth continues because of technical change, increases in human capital, increases in labour, and increases in depletable natural resources. To this is added a pure time preference rate—based upon our view of utility through time. If the long-run growth rate is 2–3 per cent, and if pure time preference is around 2 per cent, then it is conventionally argued that consumption should be skewed strongly to the present to ensure intergenerational equity.

Climate change brings this conventional approach to discounting the future into question. What if future generations inherit a much hotter climate, and what happens if a hotter climate is also a more hostile and volatile one, and sea levels rise? Though the ‘ifs’ are subject to debate and dispute, were these negative impacts to manifest themselves, it might no longer be the case that future generations are better off. Technical progress and increases in human capital might still go on making positive contributions, but they might now be offset by these negative physical impacts and, indeed, the negative consequences to social and political structures such as migration, resource disputes, and social disruption might follow.

The economic literature on climate change has focused on the discount rate and, in particular, the pure time preference rate. It has been argued that we should not discount utility through time, following Ramsey’s claim that to discount the future is to show a ‘weakness of the imagination’ (Ramsey, 1928), and in consequence the state should organize the economy so that this (much lower) social discount rate allocates resources. The market, on this view, is biased to the present, and the state should correct this bias. What is perhaps less well understood is that it implies a massive intervention in the resource allocation in the economy, and hence a much bigger role for the state. The economic borders would be shifted radically.

The alternative way of thinking about these sorts of intergenerational impacts is to focus not on the pure time preference rate—and hence on the ethical dimensions of the economic borders of the state—but rather on the growth rate. If the predictions of some climate change scientists are manifest, then growth rates will be negative. The damage will be so great as to outweigh the positive effects of new technologies and advances in human capital.

The growth rate argument is not an ethical one, and it does not require a change in the pure time preference rate—though it does lower the discount rate, since the growth bit is now much lower or negative. The role of the state is to improve the efficiency of the economy, so as to correct the market failures—and, in particular, explicitly or implicitly put prices on pollution. The territory is now much more conventional: it is about setting carbon prices, regulating emissions, and subsidizing low carbon technologies.

But how much correction is appropriate? What is the social cost of carbon, given that a cost–benefit analysis would have to incorporate damages well into the future, when people

---

7 See Stern (2007), Nordhaus (2007), and Weitzman (2007) for discussion of these discounting issues. More generally, see Helm and Hepburn (2011, ch. 2).

8 This is the approach taken in Helm (2011).

9 The conventional tools assume marginal changes. If, however, there are sharp deteriorations—non-marginal changes—these tools may not be sufficient.
will have new and different technologies? Consider, for example, carbon emissions which would have their negative impacts in 100 years’ time, and consider the possibility that by then technologies may exist to manage the atmosphere in much more comprehensive ways, and which may enable energy to be harnessed in radically different ways. By then people might have cracked photosynthesis, there may be new fast breeder nuclear reactors, and energy storage may be a pervasive technology. In such a world, the costs of dealing with today’s carbon emissions may be quite low.

The important point here is that we simply do not know, and a probabilistic approach will not get us far. The current fashion for ‘Low Carbon Roadmaps’ and ‘Energy Roadmaps’ to 2050 is a case in point: how can anything useful be said now about the configuration of energy systems in 40 years time?\(^{10}\) Consider how communications technology has revolutionized not only communications but also economies and social interactions in the last 20 years. Then imagine replicating such change to energy over 40 years—or a century. Ignorance and uncertainty are the starting point in addressing the role of the state in such intergenerational issues. A possible principle to address this uncertainty is that we ensure that we pass on to future generations assets which are at least as good as the ones we inherited. While ethics cannot be entirely excluded from this sustainability concept, by focusing on the assets, attention is directed to the capability to generate income (and hence utility) rather than its weighting between the generations. In particular, the assets-based approach to sustainability does not require the (arguably Utopian) zero time preference rate.

The sustainability criterion has many variants,\(^{11}\) and it is open to the obvious criticism that there might be good economic reasons, in trying to maximize total utility through time, to allow utility to fall in some periods if it rises by more than a compensating amount in subsequent periods. The argument is that it should not be a hard and binding constraint. But the reason these difficulties arise is that the concept is treated as a utility one and, indeed, it is a generalization of the definition of Hicksian income, and the subsequent measure of permanent income by Friedman.\(^{12}\)

Restricting the concept further, so that it is a binding constraint on assets, does not avoid these difficulties entirely: if assets are valued according to their ability to generate income, then the circularity in the definition of capital is apparent. Yet, in practice, it makes a great deal of difference to start with assets. The measure can be partial and pragmatically built up in steps. We can focus on some assets first, without being complete and comprehensive. Given the current approach is precisely wrong, steps in the right direction, however small, improve the situation. Climate is an obvious concern. Scientists suggest that the pre-industrial concentration of carbon is a good baseline. We can measure how that concentration has increased—how much the atmospheric asset has been degraded. Then, on the asset-based sustainability criterion, we can focus first on limiting further damage to the asset, and ultimately on restorative measures. Biodiversity fits into this framework too—as does infrastructure.

The asset-based approach brings the concept of depreciation to the fore. Maintaining assets intact through time is essentially the idea behind current cost accounting (CCA). In order that future generations get assets at least as good as the ones we inherited, capital maintenance is required. That can be a physical measure—so the assets are physically as good—or it can be

---

\(^{10}\) See, for example, European Commission (2010), Department of Energy and Climate Change (2010).

\(^{11}\) See Arrow et al. (2010) and World Bank (2011) for a discussion of the measurement issues.

\(^{12}\) Hicks (1939), Friedman (1957).
financial capital maintenance—so that the assets yield the same income. The difference between the two turns on technical progress and changes in relative prices.\textsuperscript{13}

It is immediately apparent how radical this asset-based sustainability criterion might be if applied to defining the economic borders of the state. For there is no comprehensive measure of the asset depreciation in current national income accounts. Indeed, GDP is primarily a production number, and it is possible to depreciate assets and claim the value of the depreciation is cash income and hence represent it as economic growth.

A cursory glance at recent economic ‘growth’ reveals precisely this effect. In Britain, North Sea oil and gas have been depleted by the current generation. There has been no offsetting depreciation charge, and therefore no asset replacement accounting. As a result the North Sea oil consuming generation has had lower taxes and higher spending than merited by the asset-based sustainability criterion. They consumed more, and saved less. Now that the oil and gas have been substantially depleted, the next generation will have to pay for imported fossil fuels. Their standard of living will fall relatively, and this is already partly reflected in the depreciation of the exchange rate. As we shall see, the current concerns about debt and deficits fit well into the same line of argument.

GDP is not an answer to the question: how well are we sustaining our assets?\textsuperscript{14} Why then is so much attention paid to GDP? The question it was primarily designed to answer is a Keynesian macroeconomic one. National income accounts developed with a view to Keynesian demand management. If the main focus is to stabilize the business cycle, and if the key issue is the level of current aggregate demand, then national income accounts arguably provide the required information. Keynesian demand management is all about production and expenditure flows, not assets, and depreciation is a matter for the medium to longer term. Keynes famously was concerned overwhelmingly with the short run. As noted above, in the long run he argued that the economic consequences for our grandchildren were bright because of the power of compound interest (and hence growth). For him, economics should be more like dentistry, managing aggregate demand, and for this he needed a measure of that demand (and aggregate supply) so that the economy could be managed to produce full employment equilibrium.

The state has a role sustaining certain assets through time and these turn out to have a number of other features in common. They tend to be public rather than private goods, and they tend to involve long-term sunk costs. Infrastructure networks are the classic example. We inherited the Victorian sewers, railways, and canals. Power stations tend to span decades, and much of the electricity networks were created in the 1920 and 1930s. The current generation has added the natural gas networks and the communication infrastructures.

Infrastructure is much more general than utility networks. Social infrastructure comprises long-lived human capital with considerable sunk costs. Education is an obvious example, as is health. The category might also be extended to encompass social and political institutions. These are all public goods, and by bringing this wider group of assets into the national accounts, a particular new focus for the economic borders of the state becomes the maintenance and enhancement of the public goods that are long lived and asset- (and therefore investment-) intensive. As public goods, they share the characteristics that the


\textsuperscript{14} It is not even a good answer to measuring wellbeing. The current fashion to attempt to measure ‘happiness’ is also a result of dissatisfaction with GDP—though whether happiness can be measured in any sensible aggregate way remains to be seen.
marginal costs tend to be low (or even zero), so that the issue arises as to how to finance the fixed and sunk costs. This, in turn, raises the question of how borrowing for investment in such public infrastructures should be set against borrowing for consumption. Current national income accounts tell us very little about how well we are meeting the sustainability criteria. Depleting the infrastructure and natural resources does not matter—and, worse, it may actually increase GDP in the short run.

By contrast, the core idea of sustainability is that that the future might not look after itself, and that the economic borders of the state should be so designed to reflect this. But to do so requires an asset-based approach to national income accounts, to which we now turn.

IV. The sustainable borders of the state—a new framework and an asset-based approach

To recap, the current economic borders of the state are the product of the twentieth century, and in particular of the reappraisal that happened in the 1930s in response to the economic crisis, and later an increasing emphasis placed on distribution and on transfers. The economic crisis starting from 2006 has provided a similar jolt to that of the 1930s, and ushered in another reassessment of the underlying rationale for the economic borders of the state and the economic theory that underpins it. The reappraisal arises in a context in which environmental intergenerational problems have been added to the primary objectives of the state. It turns out that an important element of the economic crisis—debt and leverage—is an intergenerational issue, too, alongside the environmental ones, and the hence the search for a new framework for the economic borders of the state has a common context in devising rules for intergenerational resource allocation. In this, it can build on the economic literature on sustainability and intergeneration modelling, and the specific research on pensions and pension funding.

In the previous section, the sustainability criterion was argued to be the central organizing concept, interpreted in terms of assets. The preferred definition was a physical one, while admitting that in circumstances where there is considerable technical progress, a financial maintenance concept may be more appropriate.

In making this concept operational, the starting point is to identify the core public goods that the state is required to supply—primarily from a market failure approach. In other words, we start with economic efficiency, with an intergenerational perspective, before intergenerational equity is brought into play. We will see that an efficiency-based market failure approach gets us a long way forward from current resource allocation. Once this is established, equity can be added on if necessary.

From the nineteenth century, the starting point is to identify the public goods associated with the minimum ‘night-watchman’ state. These are: defence, and law and order. These functions might be described as existential: no state can survive without them. The state has a monopoly on violence, and the state guarantees contract and property rights in the minimal state. The neoclassical analysis of markets starts with the assumption of a complete set of enforceable property rights, yet neither defence nor law and order in this wider sense are clearly defined or necessarily met by the state. Defence is relative to perceived threats and, indeed, twice in the twentieth century it could not be taken for granted. Property rights are relative, not absolute, and attenuated (not least by the state) in many ways.
As repeated attempts to provide a justification of free market systems have noted, the origins of property rights are hard to provide with moral underpinnings. In this sense, Marx’s oft quoted remark that ‘all property is theft’ has substance, and liberal theorists’ attempts to provide an ‘initial position’ as the outcome of some form of social contract—from Locke to Rawls—have not been very convincing. Neoclassical economics just takes the initial endowments as given, offering no justification. Recent experiences with trying to work out the property rights in the former Soviet Union have illustrated the complexity and moral confusions. In some cases—particularly in Russia—Marx’s observation has a very recent reflection: much property was, in effect, ‘stolen’.

As a result, the task of providing these minimal state functions remains core to the economic borders of the state. Many countries in the world currently still fail to do so. Without these functions, no state is sustainable.

In the late nineteenth century, the provision of infrastructure utilities gradually became a concern of the state (primarily at the local level). Postal services, roads, canals, railways, and water and sewerage, and then, in the twentieth century, electricity and telecommunications were public goods for which the state took increasing responsibility—either directly though ownership, or indirectly through regulation. These all presented significant and multiple market failures. The public goods element arose because the cost structures meant that the efficient output was represented by the sum of the marginal rates of substitution, since the marginal cost of an extra user was zero. Charging zero would not induce investment: hence provision would be inadequate. Extracting revenues to recover the fixed and sunk costs required some way of ‘taxing’—through user charges, the exploitation of monopoly, or through state provisions.

Government’s initial interests—especially in transport—had a significant defence rationale, but increasingly two further rationales emerged: to further economic growth and development; and to provide the basic social primary goods (of which the public health benefits of clean water and sewers were early examples). Access to markets and the cost of the network services feed through to the cost functions of every household and firm, and hence ‘good’ infrastructure has significant spillover effects to the wider economy. On the latter, the growth of democracy forced politicians to pay more attention to the welfare of the less well-off.

Now it is widely accepted that the provision of the services by these core utilities is a matter for the state to effectively guarantee. Privatization in the 1980s and 1990s changed the physical delivery of the services, but it is the state that remains the de facto guarantor. For practical purposes, the state now has the obligation to ensure that these core utility services are provided universally and in perpetuity.

How should the provision of utility infrastructure be organized? The starting point is to evaluate the assets required to deliver—and sustain—the services: the road and rail networks; the water and sewerage pipes and treatment works; the electricity and gas grids; and the communications networks. All of these are not only networks, but also systems. Unlike other assets, marginal analysis provides only limited guidance as to how they should be designed and expanded or contracted. Changing one part of the network typically affects all the rest. As a result, investment decisions need coordinating, and someone has to take a system approach to their maintenance and development. For this reason they tend to be monopolies,

---

and since private network monopolists cannot be assumed to maximize social welfare, they have to be regulated.

Regulation typically focuses on price and output—solving the monopolists’ incentives to reduce output and increase prices. The economic literature is full of sophisticated incentive models, which start with rate-of-return regulation and move towards setting prices. RPI–X is the modern approach. Regulated private utilities are provided with forward-looking, fixed-period, fixed-price contracts, with the intention of harnessing the profit motive to minimize costs.

In practice, RPI–X has turned out to be every bit as complex and intrusive as the much maligned rate-of-return regulation it was designed to replace, and in Britain the result has been to squeeze operating costs. Assets have been sweated, and there is considerable evidence to suggest that the incentives have had the desired effect on operating costs. However, the intergenerational issue is largely not about operating costs, but rather about investment and maintaining the value of the assets to deliver the services. The question is better focused on whether the infrastructure networks have been sufficiently invested to address capital maintenance and to ensure that the services available to the next generation will be at least as good as those currently enjoyed.

On this, the evidence is much more mixed. Playing around with the definition of asset lives, and bearing down on capital expenditure to hold down current prices, are both tempting strategies for regulators. Across the four main networks, there is a wide consensus that the investment requirements over the next decade are very large. Helm (2009a) estimates that in Britain these may be of the order of £500 billion for the four core infrastructures up to 2020—in excess of three times the current annual spend. Some of this capital expenditure arises because the objectives have changed (decarbonization, for example), and some because new networks are required (broadband, for example). But much is backlog: the energy networks are ageing, as are the railways and the major roads. The picture varies across developed countries, but under-investment is the norm. The position in the US is probably even worse than in Britain.

As noted in the previous section, the right way to think about these infrastructures is current cost accounting (CCA). The network services are required, for all practical purposes, in perpetuity. These services change in definition, and the CCA assets required are those best designed to deliver the services. Looking forward, the requirements are for a gradually decarbonized supply of electricity to every household and firm; road networks increasingly supported by an electric charging network; rail networks to provide longer-distance replacement for air travel; water quality in sufficient quantities to meet a rising population; sewage disposal to meet rising environmental standards; and universal high-speed broadband.

The value now of the assets is represented primarily in privatized company balance sheets—but not those of the state. In water and rail it is in current cost terms, in the energy networks it is on a historic cost basis, and in roads there is no valuation at all. In communications, the case for historic cost approaches is stronger in the context of rapid technical change, though even here it is remarkable how long copper wires have lasted.

Once the CCA values have been established, the next step is to identify for each network the capital maintenance requirement to meet the sustainability criterion. To see how far the current position deviates from the sustainable one, the current (relatively short term) capital expenditure requirements to 2020 provide a guide—some £500 billion—as mentioned above.
What these back-of-the-envelope numbers indicate is that the current population has probably not been meeting the sustainability criterion—by a considerable measure. If the national accounts are restated on the CCA basis just for these infrastructures, then the current level of sustainable consumption would be correspondingly lower (whether the capital maintenance was financed through user charges or from general taxation). This has particular relevance to the macroeconomic argument about whether the 2006 consumption levels (the pre-crisis levels) can be returned to, or whether the economic crisis requires a reduction in the standard of living.

The obvious objection is that these infrastructure assets are not state-owned, and hence are off balance sheet. This argument is reinforced by the creation of Public Finance Initiative (PFI) projects, which are treated as off balance sheet, too—and, indeed, the PFI approach was deliberately designed to be defined outside government borrowing.

Before considering which assets are public and which are private (and whether it matters), it needs to be remembered that there is no public balance sheet at all. The national income accounts were, as noted above, created for a very different purpose—for macroeconomic demand management. If the question to which the national income accounts are supposed to be an answer is now: ‘how well are we doing in meeting the sustainability criterion?’, then the ability to continue to deliver the core services depends upon having a set of assets which are sufficient for this purpose. It is irrelevant whether the assets are vested in the hands of private individuals and companies or public institutions. The issue is just: what set of assets is required to meet the functions? Whether, for example, the water pipes are formally vested in private companies is simply a reflection that the state has decided to discharge its sustainability requirement in respect of providing citizens with clean drinking water by delegating it to private companies. The duty does not go away (and, indeed, the private companies are protected by the regulator whose duties include ensuring that they can finance the functions). Similarly if the lights go out, the government will be expected to step in. This is what happened in the middle of the twentieth century: the failure of private investment, which left infrastructures not fit-for-public-purposes, led to widespread nationalization.

The sustainability criterion suggests that the schools and hospitals under the PFI are also necessary assets to discharge the state’s functions under the sustainability criterion, as are the electricity and gas networks, and the road, rail and core communications networks. They are subject to public control, irrespective of the formal ownership of the assets.

These assets should then, on this criterion, all be ‘on-balance-sheet’. They become public assets, to be set alongside the public liabilities, to provide the services. Privatization did not really magically transform the public finances. The issue is whether the assets have been properly maintained to meet the public liability of providing the services. To achieve this, capital maintenance is required to keep the asset values intact. It follows that, on a public balance sheet, there should be a claim on current revenue (tax income net of user charges) equal to the capital maintenance requirement.

Taking the core infrastructure utilities into the public accounts as assets, and recognizing the liability on the state to provide these functions, would make a radical difference to our understanding of the broader public finances and the level of sustainable consumption. What would take this to a more radical definition is to include natural resources within the national accounts.

---

16 See the paper by Joe Grice (2011, in this issue).
The rationale for this is again sustainability: we should provide the next generation with a bundle of natural resources at least as good as the ones we inherited. Valuing natural resources is not straightforward, but almost any attempt to place a value on natural assets and incorporate them in national accounts through a balance sheet approach would be better than the status quo—where the value is effectively zero (because there is no balance sheet and no additional value).

A fundamental issue to be resolved before making progress is whether man-made capital can be substituted for natural capital. In other words, the question is whether more factories, buildings, smartphones, and cars can compensate for losses of biodiversity and atmospheric damage. Some argue this is precisely what constitutes economic progress: we have accumulated more and more man-made capital by depleting fossil fuels, the land, and the atmosphere. Others take a more pessimistic view, and though they may admit that some substitution is possible, they point to critical thresholds in eco-systems, and hence that the substitution is much less than unity.

This matters greatly, since a national balance sheet should set one set of assets alongside others. But the direction of travel is the same, and we start from a zero accounting for natural asset depreciation. For Britain, an initial case might be the treatment of North Sea oil and gas, setting out a depreciation rule. The Hartwick–Solow Rule and the Hotelling principle indicate the offsetting investments required to compensate for depletion of a non-renewable resource. No such accounting took place for the North Sea: put another way, the current generation expropriated the benefits of the natural resource in higher (unsustainable) consumption. By contrast, Norway established a fund for future generations financed from its oil and gas revenues.

The challenge of accounting for the North Sea is dwarfed by that of biodiversity and the atmosphere. On biodiversity, the valuation of the assets could start with the national parks, and move on to the valuation of particular species. This only scratches the surface, but again the direction of travel would be an improvement on the zero valuation. On the atmosphere, the British government (and the European Union) have adopted unilateral carbon reduction targets. The implication is that the rate of depreciation of the atmosphere has been (much) too high, and a proxy for the scale of the damage is the investment levels necessary to rectify the perceived failures relative to the unilateral target. For example, the electricity system that would be compatible with the emissions reduction target can be compared with the actual system, and the capitalized difference between the two is equal to the under-provision (and hence over-consumption) by the current generation. In the case of electricity, roughly half the £500 billion expenditure by 2020 mentioned above could be ascribed to this shortfall.

Infrastructure utilities, PFI projects, and natural resources are the building blocks of a sustainable set of national accounts, and if these were properly accounted on a CCA basis, it is obvious that current consumption would be seen to be too high. In other words, standards of living would have to fall. This is before we turn to the national debt, the deficits, and the social liabilities for pensions, health, social care, and so on, which have been highlighted by the economic crisis.

---

17 The most recent comprehensive attempt to measure the volume of diversity is the so-called TEEB study, United Nations (2010).
18 See Hartwick (1977), Solow, (1986), and Hotelling (1931).
19 See for example, Defra (2011).
During the twentieth century, the state assumed the function of ensuring the provision of universal pensions, directly for its own expanding set of employees—and for those privately employed, it promised a safety net. For as long as the economy kept on growing, rising tax revenues would take care of the liabilities, as they would of private pension promises. Thus each generation could write a chain letter for future generations: they would be (much) wealthier, and hence could afford to pay.

Two separate factors have undermined this commitment: the population dynamics have changed as people live longer and the population ages; and the economic growth may not be sufficient. The first increases the liability; the second reduces the growth of tax revenue. Therefore, either the value of pensions has to fall (by raising the age threshold and reducing the benefits), or taxes have to rise as more of national income is shifted to paying pensions. The problem is further compounded by the increasing number of voters in retirement and the power of the public-sector interests.

From the perspective of the national accounts, pensions represent a liability and a claim on future revenues. A national balance sheet should incorporate these numbers. The recent ‘Whole of Government Accounts’ is a first step towards this. It has suggested that the public-sector pension liability should be capitalized at £1,133 billion (HM Treasury, 2011).

Health provides a further example of the misleading nature of public accounts. Health liabilities are widely argued to be on a rising trend. People live longer and medical technology offers more treatments, especially in the last years of life. As a result, the proportion of national income devoted to health is likely to rise. The question is: how much should be put aside for health maintenance, given that the cost of sustaining the function is rising? To date, the chain letter approach applies here as in pensions. The sustainable accounting approach would establish a health fund provision (and a pension fund provision, too), setting aside assets to meet the liabilities: people would make provision for their future health needs, either through private savings, and insurance, or a public fund.

Most of the focus of attention during the current economic crisis is on the national debt and the fiscal deficits. How much national debt is optimal? When and to what degree should governments run deficits (or surpluses)? Economic theory provides some guidance on the question of debt, with the concept of ‘Golden Rules’, which in practice boil down to the simple idea that borrowing for investments which pass the positive net present value (NPV) test is a good idea. The Keynesians add the idea that borrowing within the cycle might be a good idea, too, provided the net result is a balance over the cycle.

As argued in section II above, the focus in the twentieth century turned on the within-cycle borrowing and macroeconomic management. The investment bit was neglected or, worse, was exacerbated by the needs of short-term cash management. Successive fiscal crises were often met by the expedient of cutting capital expenditure (for the benefit of future voters) in the name of maintaining current spending.

It is beyond the scope of this article to address the desirability or otherwise of short-term macroeconomic management of the cycle, except to reiterate the point made in section II that the political incentives to take a risky (optimistic) view of economic growth, to redefine the length of the cycle, and to use other manipulations in order to favour short-term spending, tend to gradually ratchet up the deficits and aggregate debt.

What remains is debt. In a balance sheet approach to national accounting, following the sustainability criterion, this would be represented in national accounts as a public-sector liability. To the extent that this debt has been accumulated as a result of current spending rather than investment, it becomes a pure public liability. It is like negative capital maintenance.
It does not, of course, follow that all debt is necessarily a bad idea—if it is for investment, and if that investment does in fact pass the NPV test \emph{ex post} (and hence capture by vested interests and appraisal optimism are avoided), then it is offset by assets. But to adjudicate on this point requires a balance sheet—which is precisely what the GDP accounts do not provide.

Borrowing for investment requires financing. In current policy debates, there is a tendency to see each investment project as a discrete requirement, and therefore facing an unconstrained supply of funds. Provided the NPV is positive, then the markets will oblige. Thus a country such as Britain could continue to borrow because the value of assets created will run ahead of the liabilities. This borrowing relies upon internal or external savings.

What then is the optimal level of savings? The answer to this question is roughly: a level of savings sufficient to finance the capital maintenance charge on the infrastructure considered in the widest sense, plus the life-cycle provisions for health and pensions. There have been a number of estimates of what this might mean for the savings ratio. Excluding the wider resources, and excluding the utility infrastructures, Weale (2011) has suggested this might be of the order of 10 per cent.\textsuperscript{20} Compared with the current savings ratio, this would be a radical change.

Yet the national balance sheet approach outlined above, on the basis of the sustainability principle, suggests that even this number is probably too low. In a closed economy, with national borrowing paid out of national savings, the infrastructure investment would need to be financed from reduced consumption. As noted above, this is indeed precisely what happened after the Second World War. With capital controls, and a lack of access to (much diminished) international capital markets, consumption was suppressed to produce a surplus which the government applied to reducing the national debt and to financing the reconstruction requirements after the war. For the current generation, in the absence of capital controls, the expropriation by government of a surplus of savings from consumers would be both politically difficult and might result in capital flight. But the fall in the standard of living will nevertheless happen—it is a matter of profile and timing, not direction.

How higher savings would be translated into investment is, in part, an institutional question. This is the role of financial markets and financial institutions. In the case of the infrastructure networks, political and regulatory risk is a core issue: the fixed and sunk costs are exposed to the time inconsistency problem, since marginal costs are low. There is always the political and regulatory temptation to promise that these fixed and sunk costs will be honoured \emph{ex ante}, and then \emph{ex post} to force prices down to marginal costs. To make the commitment credible, some form of guarantee or contract is required, so that the political and regulatory risk then lies with governments and regulators, not investors.

The \emph{regulatory asset base} (RAB) is just such a commitment in the regulated utilities.\textsuperscript{21} Once capital expenditure has been completed, the assets transfer to the RAB, which is then protected by a guarantee that the utilities’ functions can be financed, thereby protecting the RAB. This guarantee is why utilities always have a public element. Financial institutions can then invest in RAB-based debt with a risk profile similar to government bonds. In effect, the RAB represents the assets required to sustain this provision of the public goods through time, with a supporting public guarantee.

\textsuperscript{20} See also Weale (2011), Barrell and Weale (2010), and McCarthy \textit{et al.} (2011).

\textsuperscript{21} On infrastructure funding, see Helm (2009b).
This approach can be generalized: in effect, pension and life funds with long-term liabilities can invest at low risk in assets whose very existence is an underpinning of the economy in the future when these liabilities arise—they literally contribute to making the economy sustainable. The current generation invests for its pensions in the assets for the next generation that will in turn help to underpin their pension provisions. It is a further step to set up the institutions that might intermediate in cases where these intergenerational assets fall outside the existing utility RABs, but have similar characteristics. For example, PFI assets could be ‘bought’ into a RAB-type arrangement once the schools and hospitals had been built. Renewable and other energy assets are another example.22

An infrastructure bank could formalize this role, more effectively allocating risk, and carrying over the RAB-type guarantee through an assurance that the functions will be financed. The infrastructure bank would therefore bring pension and life fund savings across to meet the investment requirements, creating assets to match the liabilities.

V. Conclusion: a comprehensive asset-based approach to the economic borders of the state

As the objectives of the state have widened to include not just efficiency and distributional goals, but also to sustain the economy through time, the treatment of debt and the treatment of capital expenditure has come to the fore. Current economic policy focuses on debts measured in GDP terms and debt-to-GDP ratios. These numbers tell us very little, since it is consistent with falling deficits and falling debt-to-GDP ratios if the government reduces capital expenditure in general and fails to provide for capital maintenance of the infrastructures, considered in the widest sense. There is no national balance sheet, and hence no accounting for changes in assets and liabilities.

This matters not only because we do not know what the current financial position of the government is, except in cash terms. It also matters because the national accounts give no signal as to whether we are transferring to the next generation assets as least as good as the ones we inherited. There are no generalized current cost accounts, and no charge on the public revenues for capital maintenance.

The result is a large distortion from capital expenditure to current expenditure—towards consumption now, and against investment. This is in part a consequence of a Keynesian approach to macroeconomics, which focuses on aggregate demand, without paying due attention to the difference between consumption and investment within aggregate demand. The difference from a sustainability perspective is that investment creates demand and assets, whereas consumption creates demand but does not necessarily result in new assets being created through derived demand. In a balance sheet context, this makes a very considerable difference. Borrowing to invest does not necessarily create net liabilities for future generations; borrowing for consumption may well do so. The level of borrowing is no longer in itself an interesting question: the issue is the extent of the gap between assets and liabilities.

A balance sheet approach with a full account of asset maintenance costs would radically alter the understanding of the state of the economy. Where these assets include the

22 For a description of how this might work, see Helm (2009b).
atmosphere and the biodiversity, as well as natural resources such as North Sea oil and gas, a radical rise in intertemporal savings might be required to meet the sustainability criterion. Current living standards would be revealed as high relative to the sustainable path, savings would have to go up, and then the task would be to ensure that these savings are turned into productive investments. It is in this context that there is a case for revisiting the institutional structure of the state.

These reforms to accounting, and the consequences for savings and investments would represent a significant change in the economic borders of the state. The current borders, inherited from the twentieth century, are no longer fit for sustainability purposes—just as the state which had developed in the first half of the twentieth century proved inadequate to the tasks of confronting the Great Depression and the rising demands for redistribution and welfare provision. As in the 1930s, a major economic crisis now provides a powerful basis for a reassessment. The macroeconomic policies of the last two decades have resulted in unsustainable deficits. They have also been accompanied by a lack of investment in capital maintenance and in the wider environmental assets. The balance sheet approach, based upon sustaining core assets through time, facilitated through new financial institutions, provides the basis for a new approach.

References


