Global Forum on Transport and Environment in a Globalising World
10-12 November 2008, Guadalajara, Mexico

Mind-forg'd Manacles -- The Constraints to Optimising Urban Transport Policy

Rana Roy, Maunsell/AECOM, Sydney, Australia

OECD

International Transport Forum

FOREWORD

This paper was prepared by Rana Roy, Maunsell/AECOM, Sydney, Australia, as a contribution to the OECD/ITF *Global Forum on Transport and Environment in a Globalising World* that was held 10-12 November 2008 in Guadalajara, Mexico.

The paper addresses the constraints to optimising urban transport policy. Minor modifications have been made to the paper in the aftermath of the Global Forum.

The views expressed are those of the author, and do not necessarily reflect the views of the OECD or its member countries.

TABLE OF CONTENTS

MIND-FORG'D MANACLES -- THE CONSTRAINTS TO OPTIMISING URBAN TRANSPORT POLICY

1. Introduction

1. This paper argues a definite thesis: aiming at first-best is the best guide to policy.

2. Urban transport policy, like any other field of policy, operates in an inevitably second-best world. But the constraints to developing and implementing a welfare-optimal policy in this field are, on closer examination, often imaginary ones, and where real, are often negotiable. It follows that we can, and should, move steadily toward optimising policy settings in urban transport.

3. The paper begins with an exposition and interpretation of the welfare optimum, defined in neo-classical economic theory as a heuristic device and a guide to policy, rather than as a description of the real world. In this view, a dynamic real-world economy is necessarily at variance with Walrasian equilibrium – necessarily second-best relative to the static welfare optimum at any given time. But in the absence of compelling evidence to the contrary, it is precisely the welfare optimum, correctly understood, that provides the best guide to policy. In contrast, the theory of the second-best can be a bad guide to policy.

4. Next, the paper delineates the economic content of an optimal urban transport policy, derived from the opening theoretical argument, and covering three key constituent elements: pricing, investment and regulation. This includes: pricing trips in all modes at or close to marginal social cost; investing in projects that offer positive net present values, calculated on the basis of accurate discounting; and maintaining a minimum of regulation to complement the objectives of optimal pricing and investment.

5. It then proceeds to identify and explore, in relation to each of these three constituents of policy, three sets of possible constraints: economic and financial constraints; constraints arising from politics and civil society; and governmental, including especially inter-governmental, and related institutional constraints. The paper contends that it is this last-named set that supplies the most significant constraints to developing an optimal urban transport policy – but that these, too, are negotiable.

6. Thus, the conclusion supplies a positive message to public authorities responsible for urban transport: the constraints that deter them from moving toward the optimum are, often as not, "mind-forg'd manacles" from which they can indeed break free.

2. The welfare optimum in a second-best world

7. The history of ideas offers many ironies, but here is one of the greatest. The recognition that real-world market economies are imperfect – that they do not conform to the textbook models of neo-classical economic theory and do not automatically deliver the welfare optimum described therein – has long been successfully marshalled in the service of the argument that governments should leave the market alone rather than intervene to move it closer to the optimum.

8. And here is a closely related and even greater irony. Those who claimed the mantle of realism to dismiss the advice of "textbook models" and proclaim the virtues of the unfettered "free market" have in

practice imposed a far greater extent of government intervention than is suggested by the welfare criterion in the said textbooks. The ideological climacteric represented by the electoral outcomes in the UK in 1979, the US in 1980 and several other OECD member-countries thereafter has not resulted in any serious reduction in the overall size of government, but only a change in the aims and outcomes of its interventions.

9. It is now clear that 2008 marks the beginning of a new climacteric, even if its destination is as yet unclear. For in the last three weeks alone, in the first three weeks of September 2008, the government that has been most strident in its advocacy of the free market, the current US Administration, has carried out the greatest single act of nationalisation in economic history, with its take-over of Fannie Mae and Freddie Mac, bringing over $5 trillion in assets into government ownership and adding over $5 trillion to US public debt – and followed it up two weeks later first by nationalising the world's largest insurance company, AIG, with assets of around $1 trillion, and then by announcing its commitment to develop a plan to spend around $0.7 trillion to buy up illiquid mortgage-backed assets from a range of financial institutions with significant operations in the US. Quantitatively, this dwarfs by far the extent of intervention recommended in this paper in pursuit of the first-best welfare outcome.

10. Irrespective of the detail of the emerging US plan and its impacts, the ideology of non-intervention has been decisively dethroned. Now therefore is the right moment to rehabilitate the case for welfare-optimising intervention, founded on an accurate understanding of the economic theory which underpins it.[1]

11. Neo-classical economic theory states[2] that, at any given time, and for any given set of initial endowments, an economy-wide competitive equilibrium, in which demand from utility-maximising consumers is matched by supply from profit-maximising producers in every market, will generate an optimal allocation of resources. In this hypothetical state of Walrasian equilibrium, following Walras (1877), each good will be made available for consumption at its marginal cost of production – and the marginal social benefit gained from the last unit consumed will equal the marginal social cost of the last unit produced.

12. The proximate condition that sustains this equilibrium is perfect competition. If prices are above this point, existing producers and new entrants will be motivated to expand the quantity supplied so as to pocket the difference between price and marginal cost – this expansion will continue until there is nothing left to pocket. And if prices are below this point, and some units of the quantity supplied are exacting a marginal cost higher than their price, thus putting their producers out-of-pocket, the producers in question will soon cease to supply those units. If perfect competition does not obtain – if for example, in any given market, a single firm were to acquire a monopoly and to raise prices and limit supply without fear of other firms entering the market to compete away its profits – then the market will deliver another equilibrium, one in which prices are no longer equal to marginal costs.

13. But the proximate condition for the welfare-optimality of Walrasian equilibrium, its superiority to all other equilibria, is simply the equality of price (P) and marginal cost (MC). For when prices are

---

1 I mean here the tradition of theory descending from Walras (1877) through to Arrow and Debreu (1954) and thence into the textbooks of the last half-century – for example, Samuelson and Nordhaus (1995), as well as earlier and later editions of Samuelson's Economics – and including in particular the contribution of Pigou (1912) and the closely related contributions of Lerner (1934), Hotelling (1938) and Hotelling (1939), Lange (1942) and Lerner (1944).

2 What follows borrows from and builds upon the argument developed in several of my papers over the last decade, including in particular Roy (1998), Chapter 1.3 and Chapter 2 of ECMT (2003), Roy (2006a) and Roy (2007).

either above this point or below it, *the gain to the winner* – the producer in the former case, the consumer in the latter – *is less than the loss to the rest of society*. This difference, the deadweight welfare loss, is the inevitable result of the deviation of prices from marginal costs. Thus, following Lerner (1934), we are able to write the condition of the welfare optimum in a single, simple formula: P = MC.

14. From Walras to Arrow, economists in this theoretical tradition have been well aware of the fact that real-world economies do not, and cannot, conform precisely to Walrasian equilibrium. By formalising the full set of conditions for its existence, including the existence of forward markets for all goods, Arrow and Debreu (1954) demonstrated therewith the limits of its applicability to the real world – a world of incomplete information and uncertainty. Hence, relative to the welfare optimum, all real-world markets will "fail" to some extent – if only because the attainment of this optimum in any one market requires that it be attained simultaneously in all markets, including all forward markets.

15. More importantly, there are real-world markets that do not conform even approximately to the optimum. And the main advances in welfare economics, including the P = MC rule, have been developed precisely in the course of investigating the markets that do fail in a non-trivial sense.

16. These include the markets that are the subject of this paper, the markets of the transport sector, and cover the two main cases of market failure that are addressed in this paper: the presence of externalities and therewith the divergence of *social* marginal costs and benefits from *private* marginal costs and benefits, first addressed by Pigou (1912), and the problem of pricing in industries characterised by increasing returns to scale and natural monopoly, explored by Lerner, Lange and Hotelling in the 1930s.

17. More recently, Arrow has been one of the first to address the great market failure implicit in climate change, playing a leading role in the work of the Intergovernmental Panel on Climate Change.

18. And last but not least, in continuity with the major economists in the preceding classical tradition, Walras also recognised the peculiarity of the markets for land. The ownership of scarce land and residential and non-residential property built on it – as well as scarce natural resources, including especially oil and gas – generates economic rents unrelated to the cost of production. Accordingly, successive economists in the Walrasian tradition have continued to identify the taxation of economic rents from privately-owned land and other natural resources as a non-distorting, and hence appropriate, form of taxation – and an efficacious means of reducing reliance on distorting, and hence welfare-reducing, taxes.[3]

19. The theoreticians of the welfare optimum are the last people who need to be reminded that real-world markets do not conform to our textbook models.

20. To interpret and criticise Walrasian equilibrium as a description of the real world is thus to mis-describe radically what it is and what it seeks to achieve. Rather, it is best understood as a heuristic device, the starting point of a research programme and a programme of policy development and implementation. Thus understood, it can discover and specify for science the meaning and measure of social welfare and the circumstances in which markets, unaided, cannot deliver the welfare optimum – and can discover and

[3] As indeed does ECMT (2003). It may be noted that taxation of privately-appropriated rents is not the only means of achieving this outcome. Walras, at least initially, recommended the nationalisation of land and the use of rents to fund public services – a programme that successive governments in Hong Kong, both under British rule and following the re-union with China, have carried out with conspicuous success. Some oil-producing countries have successfully collected the economic rents from oil by means of royalties, others by means of sovereign ownership. But given the prevailing pattern in OECD member-countries, this paper follows ECMT (2003) in assuming taxation rather than nationalisation as the preferred instrument of choice.

specify for governments why and when and how to intervene in order to move markets closer to that optimum.

21. The policy prescription that follows from this research programme is in fact reducible to a remarkably simple and limited set of interventions, notwithstanding the complexity of the detail of implementation. It includes pre-eminently:

- Competition policy to re-enforce competition in all actually and potentially contestable markets;
- The targeted application of externality taxes and other related instruments to "internalise" and reduce the most significant externalities across all sectors;
- A comprehensive tax-and-subsidy policy to correct prices in those markets characterised by systematic and significant market failure, including the markets of the transport sector.

22. Importantly, the guiding principle of policy here is strictly consistent, irrespective of the varying extent of intervention that may be required: that is, to aim at the application of the P = MC rule.[4]

23. The subsequent sections of the paper successively spell out the application of this rule to the transport sector, identify the constraints faced by decision-makers seeking to apply the rule and indicate the ways and means of overcoming these constraints. But first, it remains to complete the argument at a more general level and in relation to the economy as a whole: that aiming at this first-best rule remains the best guide to policy despite the caveats that limit its applicability to real-world economies.

24. The welfare optimum that follows from the P = MC rule is a static optimum: it applies to a given point in time and for a given set of initial endowments. Even if it were possible to apply it absolutely – which it is not, thanks to incomplete information and uncertainty if nothing else – its absolute application in, say, 1900 would not have made the world better off in 1900 than it is today. This is because, over time, human society is capable of expanding its knowledge of nature and applying this knowledge to its production processes, thereby increasing the effective sum of resources available to it. And over the last century, human society has demonstrated that capability more spectacularly than in any previous century. Of course, the same observation was famously made in 1848 in regard to the preceding one hundred years. And it will surely be made again in 2050 in regard to the one hundred years since 1950.

25. Now the "phenomenon of development", as Schumpeter (1911) observed, is a disturbance of equilibrium. Innovation – for example, the introduction of power-looms in an industry operating with hand-looms – enables the innovating firm to lower its own costs far below the prevailing norm. Eventually, the innovation is diffused through the industry so as to create a new and lower norm of costs and prices, enabling consumers to enjoy the fruits of innovation as a welfare gain. But since it takes a more or less lengthy period of time for diffusion to do its work, the first fruits belong to the innovating firm, which can exploit the gap between its own costs and the prevailing prices to earn a "super-normal profit". It enjoys, in effect, a temporary monopoly. And this monopoly is the seed of innovation, the incentive that brings forth the innovative effort, from which society as a whole ultimately benefits.

[4] As noted earlier, a taxation policy aimed at maximising revenues from non-distorting taxes – as distinct from a taxation policy aimed at correcting price distortions – is an important additional policy programme that follows from the criterion of welfare-optimisation. But this topic *per se* lies beyond the scope of the present paper. It enters into the discussion here indirectly, and mainly in order to refute the oft-repeated, but groundless, suggestion that a policy aimed at correcting price distortions in particular markets needs recourse to distorting taxes elsewhere in the economy.

26. Moreover, for Schumpeter (1942), insofar as the innovating firm is able to extend its temporary monopoly through restrictive practices, the greater is the incentive for innovation and the greater the likelihood of its occurrence. Hence, Schumpeter (1942) concludes: "A system which is efficient in the static sense at every point of time can be inferior to a system which is never efficient in this sense; because the reason for its static inefficiency can be the driver for its long-term performance."

27. Schumpeter's (1911) original insight is quite true. Insofar as it is impossible to introduce and diffuse innovations and to change all corresponding prices instantaneously, a dynamic real-world economy is necessarily at variance with Walrasian equilibrium. In this sense, it is necessarily "second-best" relative to the static welfare optimum at any given time. And accepting this particular deviation from the optimum is indeed a corollary of achieving ever-higher levels of welfare over time.

28. What does *not* follow in the least is the proposition that governments should abandon the P = MC rule as a guiding principle of economic policy – in the general case, by means of competition policy – and accept the principle of non-intervention in markets, up to and including accepting restrictive practices by firms seeking to hold and prolong temporary monopolies.

29. On the contrary, an active and aggressive competition policy is the best means of securing the fruits of Schumpeterian innovation. This is not only because competition policy facilitates the process of diffusion and thereby speeds up the transformation of the gains of innovation into gains in welfare. It is also, and primarily, because, as is well described in Schumpeter (1942), productivity-enhancing innovation is only one among many avenues by which to gain and hold a monopoly position and corresponding profits. Hence, by closing down as many other avenues as possible and obliging firms to search for innovation as the main source of "supernormal profits" – in modern terminology, to search for technological rents as the main source of rent – governments can provide a greater incentive for innovation than is provided by the tolerance of restrictive practices at the expense of consumers. Thus, *we can advance the pace of innovation at the same time as advancing the pace of diffusion*.

30. The record of the last half-century offers compelling evidence on this point. In most OECD member-countries, active competition policy has eliminated many types of restrictive practices and increased the intensity of competition in most contestable markets. The result has been to push most such markets to operating closer to the P = MC desideratum than would otherwise have obtained.

31. In the UK, for example, and before the onset of the current recession, prices across *all* sectors were, on average, 20% above marginal costs according to the best estimates available.[5] Given that this average is, in large part, the result of significant monopoly power exerted in a small number of a-typical (and typically volatile) markets – oil and gas, land and property, finance and insurance – the deviation of prices from marginal costs in most markets is in fact remarkably small.

32. Despite this, the pace of innovation and productivity growth in the last half-century across all sectors of production has been greater than in any previous age.

33. In short, recognising the positive role of Schumpeterian innovation does not licence a Schumpeterian defence of monopoly. Least of all does it justify a defence of monopoly where the aim of the monopolist is not the collection of technological rents by means of slowing down the pace of diffusion of any particular innovation, but simply the collection of rent *tout court*.

34. Indeed, the question that is posed in the light of the tumultuous events of 2008 is not how to restore monopolistic privileges across all sectors in order to promote innovation, but rather how best to

[5] See DfT (2006), in particular, Annex 1, "Theory and Evidence", and the studies cited therein.

close down, or transfer to society as a whole, the rents available in those sectors where monopolistic privileges have not been effectively tackled: as detailed above, oil and gas, land and property, finance and insurance.[6]

35. And this in turn illustrates a more general point: in the absence of compelling evidence to the contrary, the welfare optimum, correctly understood as a heuristic device that can help to discover and specify why and when and how to intervene in markets, remains the best guide to policy.

36. In a similar vein, the true insight in Lipsey and Lancaster's (1956-57) seminal work on the "general theory of second best" should not license the abandonment of the first-best as a guide to policy.

37. The true insight is this: if there is a constraint that prevents the application of first-best values in some part of the system but not in others, then there will be a set of correct departures from first-best values in the unconstrained parts of the system that will deliver a better result than will the application of the theoretically first-best values in the unconstrained parts.

38. To put the point less formally: relative prices matter. If a particular price is wrong and cannot be corrected, then selecting a correct set of departures in respect of the other relevant prices in order to offset the original distortion will deliver the best possible outcome under this constraint.

39. However, as is argued below in relation to the problem under discussion, there are several grounds for supposing that second-best theorising can be a bad guide to policy. And it would be as well to summarise here the gist of the argument at a more general level rather than in a sector-specific context.

40. First, it should be noted that second-best prices are of course just as "theoretical" as first-best ones. In practice, it may be enormously difficult and costly to discover and implement these optimal departures from the P = MC rule. There is therefore a serious possibility of getting it wrong – that is, selecting sub-optimal departures that deliver a worse welfare result than simply applying the rule regardless.

41. More fundamentally, it is necessary to investigate the nature and origin of the constraints that supposedly necessitate the search for these optimal departures. And insofar as these constraints are themselves the result of policy decisions – decisions that reflect and signal a relative indifference to the welfare result on the part of the relevant decision-makers – it seems highly improbable that another, and possibly subordinate, part of the decision-making system will be able to implement and sustain such complex corrections to offset the effects of the original policy decisions.

42. Finally, it needs to be stressed that whilst the theory can sometimes justify departures from first-best values in order to offset the effects of market failure, it certainly does not justify the use to which it has sometimes been put: departures that compound market failures.

43. To sum up: we know that Walrasian equilibrium will not be realised perfectly in this second-best world. We know that the P = MC rule will not deliver precisely the welfare optimum that would obtain in the event of its perfect realisation. But it does not follow that there is a feasible alternative that can deliver a better welfare outcome than a policy aiming at the application of the P = MC rule on an economy-wide basis – by means of competition policy in most markets, internalisation of the most significant externalities in all markets, and a comprehensive tax-and-subsidy policy where market failure is systematic and significant.

---

[6] This is the question now being debated with vigour in the pages of the *Financial Times*; see for example Brittan (2008), Caldwell (2008) and FT Editorial Comment (2008).

44. Therefore, in approaching this second-best world, this paper follows a different path to that indicated by second-best theorising. Instead of turning away from the first-best and searching for the ways and means of "optimising under constraint", it interrogates the actual and imagined constraints to a first-best policy, searching rather for the ways and means of overcoming the relevant constraints.

3. Defining optimal policy in urban transport

3.1 Pricing

45. In defining an optimal policy for urban transport, it is best to begin by specifying the key elements of optimal policy for inland transport in general. For the former needs to include each of these elements as well as others that are specific to the urban context.

46. The markets of the transport sector are indeed characterised by systematic and significant failure. Two types of market failure obtain, compounding rather than offsetting each other.[7]

- As a function of its technology, much of modern transport infrastructure is characterised by increasing returns to scale. The fixed costs of production are often high; marginal costs are often low. The result is that the average cost of production can be considerably greater than its marginal cost. This is most pronounced in the case of rail, including metros. Hence, *ceteris paribus*, rail services would need to be priced far above marginal cost in order to cover the total cost of production – that is, depreciation of past investments, maintenance and operation.
- The use of transport infrastructure generates large externalities. These include local and global pollution, accidents, and the external costs of congestion imposed by new users on existing users whenever the infrastructure is operating at or above capacity. This is most pronounced in the case of urban road use, least so in modern metros. Hence, road use would be priced far below its marginal social cost if it were priced only to cover the cost to road authorities of depreciation, maintenance and operation.
- And insofar as the modes of transport are substitutes: *ceteris paribus*, their relative prices would prompt a welfare-reducing substitution between modes, compounding the welfare loss in each.

47. In order to achieve the welfare optimum, it follows that:

- a subsidy will be required to cover fixed costs so as to reduce the marginal price faced by the user to the marginal social cost imposed by his use – and thus price in welfare-increasing consumption that would otherwise be priced out;
- and a tax will be required to internalise external costs so as to raise the marginal price faced by the user to the marginal social cost imposed by his use – and thus price out welfare-reducing consumption that would otherwise be priced in.[8]

---

7 The following summary borrows from the text of my published papers cited in n. 2, including in particular Roy (2006a).

8 Arguably, a transfer to cover fixed costs should not be described as a "subsidy" since what obtains in its absence – prices far above marginal social cost – is best described as an excise tax, as it was described in Hotelling (1938). And a user-charge to internalise external costs should not be described as a "tax" since what obtains in its absence – prices far below marginal social cost – is indeed a subsidy, as is increasingly recognised not only in economic literature, but also in debates on formal classification, such as in OECD (2005). *Pro tempore*, however, we continue to follow the conventional terminology so as to avoid any confusion.

48. As indicated above, the market failure in transport is likely to be most pronounced in urban transport. This is primarily because the external costs of congestion imposed by private vehicles are likely to be most pronounced in urban areas – in the UK, for example, Eddington (2006) reports that almost 90% of the time lost in congestion is in urban areas. But the same also applies to several categories of local pollution.

49. Moreover, in a mirror image of the external cost of congestion, there is also an external benefit created by new users that is most pronounced in urban public transport – rail, metro and buses – known as the Mohring effect, following Mohring (1972). When the frequency of scheduled public transport services is increased in response to an increase in demand, waiting times fall for existing users. As Glaister (1994) describes it – "the new users thereby create a benefit external to themselves but internal to the system." Hence, a subsidy to price in this element of welfare-increasing consumption is appropriate.

50. Thus, in urban transport more so than in the transport sector as a whole, markets, unaided, would price rail and other public transport far above the optimal level and private road transport far below it.

51. Of course, governments across the world have long been cognisant of this double market failure in the transport sector. Consequently, they have intervened in transport markets with a mix of subsidies and taxes – in particular, taxes on fuel – in an effort to achieve an approximate degree of correction. The problem is that the current mix does not come close to aligning prices to marginal social costs. What is required now is not intervention *per se* but intervention to move the markets of the transport sector toward the optimum.

52. An extensive research effort over the last two decades has now succeeded in quantifying – not in needless decimal points, but in the orders of magnitude required to trigger the process of policy reform – the extent of outstanding market failure, and, with it, the gains from correcting it.

53. In the case of rail, the evidence from econometric studies shows that the ratio of marginal costs to average costs for passenger and freight rail is generally in the range of 0.6 to 0.7[9] Pricing to achieve full cost recovery without recourse to subsidy would thus require over-charging rail users by up to 67%.

54. Fortunately, at least in Europe, there is sufficient subsidy in place to ensure that the price of rail use is generally well below the full cost recovery price – and often slightly below rather than above the marginal social cost price. This also holds true for the UK, notwithstanding a widespread perception to the contrary. In a benchmark study for the UK Department for Transport, Sansom, Nash, Mackie, and Shires (2001) found that the ratio of revenues to marginal social costs in passenger and freight rail services stood at 0.85 and 1.13, respectively, in 1998. The evidence published by DfT in its latest White Paper on Rail[10] found the same indicator to be in the range of 0.7 to 0.8 for passenger rail in 2004.

55. The problem for rail – and, more importantly, for society as a whole – is that the price of road use is nowhere near its marginal social cost. In 1998, the ratio of revenues to marginal social costs for *all* road transport (passenger cars, trucks, buses), as found in Sansom, Nash, Mackie, and Shires (2001), was in the range of 0.36-0.5. Subsequently, in 2004, according to the evidence published in DfT (2007b), the ratio of revenues to marginal social costs for passenger cars stood at no more than 0.15!

56. *Relative prices are thus badly distorted* – principally because of the remarkable under-pricing of car traffic. And the same holds true not only for the UK, but also across Europe – and not only for Europe

---

[9] See the evidence presented in Roy (1998) and the econometric studies cited therein, from Caves, Christensen, Tretheway and Windle (1985) to Kessides and Willig (1995).

[10] See DfT (2007a) and DfT (2007b).

but also across the OECD world, and *a fortiori* so in those countries where taxes on fuel have been set at relatively low levels compared to Europe, including especially the United States.

57. Given the extent of outstanding market failure, there are large gains available from correcting it. Perhaps the most comprehensive cross-country evidence on the matter was that provided in Chapter 2 of ECMT (2003) and pertaining to the year 2000. Presented to European Transport Ministers in 2003, it helped to trigger the process of policy reform in Europe.[11]

58. For the purpose of the present discussion, there are at least three points worthy of note.[12]

59. First, despite the differences between the countries under study, the gains in each country from optimising prices are large, and similarly large, both in the aggregate and in regard to each of the major externalities.

60. Thus, in Britain, France and Germany, respectively, the overall gain in welfare was estimated at €13billion[13], €10 billion and €9 billion per year. The reduction in air pollution and CO_2 emission costs was estimated at 54%, 50% and 37%. And the de-congestion effect, as indicated by the average increase in peak-hour road traffic speed in the main metropolitan centres, was estimated at 11%, 9% and 15%.

61. Secondly, and precisely because of the greater extent of the deviation of prices from marginal social costs in urban areas, the gains from correcting it accrue overwhelmingly to urban areas.

62. Thus, in Britain, France and Germany, respectively, the gain in welfare in urban areas, when measured as a percentage of the welfare gain nationally, was estimated at 88%, 73% and 71%.

63. Thirdly, the optimisation of prices entails a significant change in the modal split of traffic volumes in urban areas: thus, the estimated level of optimal urban traffic volumes, measured in daily passenger kilometres, in London, the Ile-de-France and Munich-Hamburg, respectively, entailed a reduction in car traffic by 20%, 16% and 28%, and an increase in rail/metro patronage by 21%, 22% and 24%.

64. Given the evidence on the extent of outstanding market failure and on the measure of the gains available from correcting it – and given the implications of this evidence in regard to the relative utility and investment needs of the various modes of transport – it is unsurprising that governments across the OECD world are indeed committing themselves to a programme of reforming transport prices.

65. The commitment to policy reform is most explicit in Europe – as is the pace of progress on the ground.[14]

---

11 Subsequently, more detailed studies for national governments, modelling against future reference scenarios with uncorrected prices, have shown even larger gains from optimisation, building a powerful momentum in support of reform: see in particular Eddington (2006).

12 For a fuller explanation of the modelled results reported here, and of the modelling informing it, see ECMT (2003).

13 More accurately, £11 billion, which converted at the then prevailing exchange rate, was reported as €17 billion in ECMT (2003). At today's exchange rate, £11 billion converts to €13 billion.

14 That is, apart from the city-state of Singapore, which has long provided a successful example of comprehensive road-user pricing.

66. Here, as early as 1998, the European Commission had recommended the principle of marginal social cost pricing as a guide to transport pricing policy.[15] In 2003, the Transport Ministers of the ECMT member-countries endorsed the long-term aim of instituting marginal social cost pricing and agreed that near-term changes "should always move in the direction of improving efficiency."[16] In 2004, the UK Government announced in its White Paper on *The Future of Transport* its intention to commence preparatory work for a future national road pricing system[17] – an intention confirmed, albeit modified in respect of timetables and transition paths, in its 2007 White Paper.[18] In 2005, the Netherlands Government announced a provisional decision to introduce nation-wide kilometre charging for road use[19] – a decision confirmed in 2007.

67. On the ground, local congestion charging systems are now operating successfully in several cities across Europe, including London (since 2003)[20] and Stockholm (since 2006).[21] So, too, is the kilometre charging system for trucks in Switzerland[22] – and several others that have followed in its wake.

68. But progress on the ground is also evident elsewhere, including the United States. At much the same time as London witnessed the introduction of the Central London congestion charge, New York witnessed an effective increase and differentiation of the charges on the tolled crossings into Manhattan[23] – and whilst proposals for a more comprehensive pricing scheme in New York itself have yet to gain acceptance, the last half-decade has seen the launch of several successful pricing initiatives elsewhere in the country.

69. Importantly, these steps have sent a powerful signal that has helped to trigger the development of the technology required for an accurate and highly differentiated form of kilometre charging, differentiated by route, time of day and type of vehicle. For this is the new instrument that is required to tackle the full spectrum of externalities even as the taxation of fuel remains the best instrument to tackle CO_2 emissions.

70. This is not to suggest that the development of technology will permit a perfect alignment of price to marginal social cost for each and every trip – or indeed that it would justify the effort required to enforce such perfection. But that is not the point of the exercise: as theory informs us, perfect prices in any one market will not deliver the perfect welfare outcome unless complemented by perfect prices in every

---

15 See EC (1998). In preceding and subsequent EC publications – including EC (1995) and EC (2006) - the principle was more heavily qualified by cost recovery reconsiderations. But its latest communication – EC (2008a), along with the companion report, EC (2008b) – represents a return to the point of departure marked in EC (1998).

16 ECMT (2003).

17 See DfT (2004a) and its companion report, DfT (2004b), the feasibility study on national road pricing.

18 DfT (2007c).

19 See Werther (2006).

20 See in particular TfL (2004), TfL (2005) and TfL (2006). The results reported in more recent monitoring reports show a measure of deterioration in the de-congestion effect, mainly as a result of the take-over of Central London road space by a very large-scale maintenance works programme initiated by Thames Water.

21 See Stockholm Stad (2006).

22 See ARE (2004).

23 See Zupan & Perrotta (2003) for a summary of the state of play in New York at the time – with 22% of vehicles coming into Manhattan already paying a toll, that is to say, roughly double the number of vehicles paying the congestion charge to come into Central London – and an exposition of alternative scenarios for the further development of congestion charging.

market. The point rather is that, whereas today the deviation of prices from marginal social costs in this sector is greater by orders of magnitude than what obtains on average across all sectors,[24] it is now possible to aim at reducing that deviation to a minimum, and indeed to less than what obtains across all sectors.

71. Hence, the optimal pricing policy that follows from the theory and evidence presented above is best described as it was described in the introduction: a policy of *pricing at or close to marginal social cost*.

3.2 Investment

72. The discussion above, including all the results reported from the studies cited, limited itself to the question of how best to price the use of existing infrastructure. An optimal policy for the transport sector must also address the question of how best to adjust capacity in infrastructure in response to changes in demand.

73. There are two preliminary points to note here. First, pricing and investment do need to be treated separately, though both are essential for the long-run performance of the system. The technological characteristics of transport infrastructure – in particular, the indivisibilities or "lumpiness" of the investment required for new infrastructure and the long lead times required to put it in place – are such that the most reliable basis for calculating optimal prices are *short-run* marginal social costs, the costs that arise from the use of existing infrastructure, assuming that infrastructure to be given.

74. Hence, the recommendation both in EC (1998) and ECMT (2003) to complement optimal pricing, based on short-run marginal social costs, with a programme of optimal investment, based on comprehensive appraisals of the social costs and benefits of proposed investments.

75. Secondly, there is an important sense in which the question of pricing is logically prior to the question of investment – the wrong answer to the former will generate wrong answers to the latter, quite possibly at an enormous cost to the welfare of future generations. The schedule of demand that follows from the wrong set of prices is not the same as the schedule of demand that would follow from an optimised set of prices. *Prima facie*, a correction to prices is therefore likely to alter the composition and location, the scale and the timing, of the investment required to meet future demand. But decisions are being made today on investments in infrastructural assets with an economic life of up to 100 years. Hence, assuming that a correction to the present pattern of distorted prices will indeed be achieved in less than 100 years, these investment decisions need to take account of that correction if we are to avoid denying future generations the assets they will need and bequeathing them assets that they will not need. [25]

76. On a national scale, studies in the Netherlands[26] and the UK[27] have shown that road pricing is likely to reduce significantly the future gains from providing additional road capacity. For the UK, Eddington (2006) suggests that an optimised national road pricing scheme would reduce the benefits from providing additional inter-urban road capacity beyond 2015 by 80%.

77. In short: to get investment right, we need to get prices right.

---

24 Cf. the transport-sector results cited above with the cross-sectoral evidence in DfT (2006) cited earlier.

25 For a fuller statement of the argument, see Roy (2005).

26 See CE Delft (2002) and the further analysis of these results reported in ECMT (2003).

27 See Eddington (2006).

78. That said, the guiding principle of an optimal investment policy can be stated quite simply: proceed with investment if, and only if, the investment in question offers a positive net present value at the chosen discount rate, such that the present value of its discounted future streams of benefits exceeds the present value of its discounted future streams of costs.

79. To spell it out formally, proceed if and only if:

$$NPV = PVb - PVc = b_0 - c_0 + \frac{b_1 - c_1}{1 + r} + \frac{b_2 - c_2}{(1 + r)^2} + \ldots + \frac{b_n - c_n}{(1 + r)^n} > 0$$

where NPV is net present value, PVb is present value of benefits, PVc is present value of costs, r is the discount rate, n is the final year of evaluation.

80. Clearly, in order to arrive at an accurate result, it is important to *count* all relevant costs and benefits as accurately as possible. But the choice of the rate at which these annual streams of costs and benefits are *discounted* over time is also critical – as is the choice of the evaluation period, the point at which counting stops. In fact, the available evidence suggests that recent and proposed corrections to the discount rate are likely to be of greater quantitative significance than recent and proposed corrections to the methodology of calculating the relevant costs and benefits.[28]

81. The basic theoretical justification for discounting is that society places a higher value on present consumption than on future consumption.[29] But the discounting that has often been applied in practice implies a zero or near-zero value for future consumption. Thus, by Year 30, a unit of net benefit keeps a present value of 0.3563 if discounted at 3.5% – but only 0.1741 if discounted at 6%, and as little as 0.0573 if discounted at 10%. And if we stop counting at Year 30, its value becomes, by definition, zero.

82. Of course, the point applies to all investments, but it applies *a fortiori* to the long-lived infrastructural assets of the transport sector. A 10% discount rate applied over a 30-year evaluation period is, in effect, a refusal to count the *greater part* of the benefits of infrastructure projects.

83. The same applies to long-term costs, including especially environmental costs that might impact on society over generations, climate change being the most important topical example. A 10% discount rate applied over a 30-year evaluation period might well fail to count *any part* of certain major environmental costs.

84. In short, the effect of high discount rates and short evaluation periods is to discount the future away.

85. And this is just what obtained in the last quarter of the twentieth century across much of the OECD world – and continues to obtain in several OECD member-countries today. The public-sector discount rate climbed to 10% in the US, the UK, Canada and Australia, and close to 10% in several Continental European countries (with the notable exception of Germany). And counting stopped at Year 30. The 10% discount rate was also exported to much of the developing world *via* the World Bank. And

---

28 See Roy (2006b) for a fuller statement of the argument and the evidence available on projects in the UK.

29 To be sure, there is not a universal consensus on this subject. For a textbook exposition of the theoretical controversies in the choice of discount rate, see *inter alia* Layard and Glaister (Eds.) (1994) and Spackman (2001). For a comprehensive survey of the literature on the subject from the 1920s to the present day, see Spackman (2004). Finally, for the renewed date on discount rates in the context of the valuation of the costs of climate change, see in particular Stern (2006) and Sterner and Persson (2007).

even after the downward revisions effected in several countries in the 1990s, the discount rate applied to infrastructure investments remained high: 10% in Canada, 7% in the US and Australia, and, until 2003, 6% in the UK.

86. Thus, at the turn of the century, ECMT (2001), was obliged to conclude that the result of established policy has been to "distort the relative merits of different schemes … because … a higher discount rate will tend to favour schemes which have more immediate benefits and/or more distant costs" – that is, to penalise those projects which yield the longest streams of benefits and the lowest future costs.

87. More recently, however, the UK has provided an example of how to conduct accurate discounting – and, with it, important evidence on the impact of such a correction on the evaluation of infrastructure projects.

88. The UK Treasury's 2003 Green Book established a "social time preference rate" derived from first principles as the basis for the public-sector discount rate – whilst treating other legitimate but separate concerns and objectives with other instruments. On this basis, it instituted a new standard starting discount rate of 3.5% over a 30-year evaluation period.[30] It also recommended extending the evaluation period beyond 30 years to count long-term impacts where appropriate – and a declining schedule of discount rates for these long-term impacts, starting at 3% from Year 30. Thus, in any cost-benefit analysis conducted today, a unit of net benefit in Year 30 will keep twice the value that it did at the 6% rate applicable before 2003 – and, where appropriate, will continue to be counted out to Year 100 and beyond.

89. The new discounting regime has transformed the outlook for investment in transport infrastructures. In particular, it has helped to establish the high net present values available from projects with the longest streams of benefits and lowest future costs, of which London's planned Cross-Rail project is the best example. At the same time, it has also helped to shed light on the devaluation of comparable projects in past appraisals – best illustrated in the re-appraisal of the Jubilee Line Extension to London's Underground Network, which was born against the odds after "failing" its original appraisal in 1991.[31]

90. The UK has also been at the forefront of efforts to extend the boundaries of cost-benefit analysis in order to count more accurately the impact of transport investments on the wider economy.[32]

91. Over the decades, cost-benefit analysis has developed a more or less robust methodology for calculating the *direct impact* of transport investments – whilst recognising that benefits and costs can be transferred from one party to others as the direct impact is absorbed into the wider economy. Thus, to follow the choice of examples in DfT (2006), time-savings for firms can be transformed into price reductions for consumers, time-savings for commuters can be transformed into rent increases for landowners, and so on. The new research seeks to identify and quantify the *additional impact* of transport investments on the wider economy, over and above the direct impact measured in conventional CBA– not

30 See HMT (2003). An alternative approach is to base the discount rate on government's long-term cost of borrowing. But as is shown in Roy (2006b), by either method we arrive at much the same result. The appropriate discount rate for the UK falls within the range of 3-5% - and indeed at the lower end of this range, given current projections of per capita GDP growth and current yields on long-dated gilts.

31 For a fuller discussion of these and other examples, see Roy (2006b) and the studies cited therein.

32 The key document is DfT (2006). But the relevant research effort stretches back to at least SACTRA (1999) and includes several of the papers published in ECMT (2001). For my summary of the debate as at 2006, see Roy (2006b), from which much of the present discussion is drawn. See also Eddington (2006) which reports further results for the UK building on the work of DfT (2006) – and OECD/ITF (2008) which reports and analyses a range of findings from international research in the field.

simply the qualitative change in the incidence of the impact over time but, rather, the quantitative change in its value.

92. For the purpose of the present discussion, the key relevant findings are these.

93. First, transport investments *do* generate additional impacts – even if these are difficult to quantify and even if any quantified results are modest. For it is only in the hypothetical condition of perfect competition, with prices equal to marginal costs both in the markets of the transport sector and in the markets of the transport-using sector, that the value of the additional impact would be zero.[33]

94. Secondly, additional impacts can run in both directions: they can add to, as well as subtract from, the value of the direct impact, depending on the extent of the price distortions in the relevant transport and transport-using sectors. But *if* the appraisal of the transport investment corrects for price distortions in the transport sector, as indeed it should do, then the available evidence on prices in the transport-using sectors suggests that, in most cases, the additional impact of worthwhile transport investments ought to be positive.[34]

95. Thirdly, there is a limited range of such positive additional impacts for which the current state of research can now report results that are methodologically sound and quantitatively large. Of these, the most important is the contribution of transport improvements to agglomeration economies.

96. Agglomeration economies refer to the positive externalities generated by clusters of economic activity, whereby the productivity of each firm in a given cluster is enhanced by its proximity to other firms in the same cluster – thanks to the opportunities for greater sharing of formal and informal knowledge, the access to more suppliers and larger labour markets, and so on. Now whilst each firm may be expected to take into account its own benefits in making its decisions on location, it will not take into account the benefits it creates for other firms in choosing to remain in, or relocate to, a given cluster. Hence, by improving accessibility within a given cluster or facilitating relocation to it, transport improvements can increase the effective density of employment in a given area and therewith its productivity.

97. As Eddington (2006) reports, it is in "large, high-productivity urban areas" that this effect is likely to be greatest. In the UK context, London is of course the pre-eminent site of agglomeration economies – and it is London's major rail/metro projects, from the Jubilee Line Extension to Cross-Rail, that have provided the pre-eminent examples of the contribution of new investments to agglomeration economies.

98. To be sure, it is not possible to generalise from these UK results in order to indicate a general pattern of outcomes from an optimal investment policy. In the case of pricing, there is a general pattern to the problem itself: the under-pricing and over-use of private road transport in urban areas. The analysis of the problem, when coupled with modelled results from national as well as European cross-country studies, does permit us to indicate a general pattern to the results that may be expected from optimising prices in all comparable countries: large overall gains in welfare, the concentration of these gains in urban areas, a change in the modal split of passenger transport in our cities, with a reduction in car traffic and an increase in rail/metro patronage. But investment needs will vary widely across countries in line with the varying pattern of past investment decisions and hence the present endowment of infrastructure. Nor has any one country attempted to model the results of an optimised appraisal of all feasible investment projects.

---

[33] See SACTRA (1999) and the discussion thereof in Vickerman and Monnet (2001) and Roy (2001).

[34] As argued in Roy (2001), DfT (2006) and Roy (2006b).

99. Nonetheless, there are at least some broad conclusions that may be drawn in regard to urban transport. Accurate discounting will favour projects with the longest streams of benefits and lowest future costs, making projects with large initial funding requirements more viable, including initially expensive road and rail/metro schemes in urban centres. So too will the new recognition of agglomeration economies. But the opportunity costs of providing additional road capacity in urban centres will remain high, sometimes forbiddingly so, given the high value of its alternative uses in residential and commercial property as well as in public realm space. And correcting for prices will reduce the demand for, and the benefits from, additions to urban road space relative to what would otherwise obtain. To the extent that the effect of these factors is not offset by other factors, an optimal investment policy is therefore likely to entail *inter alia* an expansion in rail/metro capacity in many urban centres across the OECD world, and not only in London.

3.3 Regulation

100. The welfare-optimal pricing and investment rules articulated above are interventions that correct the market but do not replace it. Thus corrected, the market mechanism continues to operate as it should. It only remains to add that an optimal urban transport policy also requires a non-market mechanism as its third constituent: a minimum of regulation to complement the objectives of pricing and investment.

101. Of course, the transport sector is already regulated tightly in a number of areas, in particular those relating to safety. Safety standards for vehicles in all modes, the rule book of the railways, road traffic management, drivers' licensing, compulsory insurance, policing and incidents management – all these are a familiar part of the landscape.

102. They also provide an important guide to the regulation required in other areas. For each of these regulatory measures are intended to achieve outcomes that market mechanisms might well achieve but could do so far less efficiently.

103. For example, if vehicle safety standards were not in place and vehicle producers were legally entitled to produce less safe vehicles at a lesser price, the exercise of consumer choice would, over time, lead to producers of safer vehicles putting the producers of less safe vehicles out of business – but only after needless additional fatalities, injuries, associated costs and consequent losses in welfare.[35]

104. The same principle applies to the regulation relevant to the present discussion: the regulation required to complement the objectives of optimal pricing and investment. Here, too, regulatory measures are preferred to market mechanisms if they are a more efficient way of achieving the same objectives and precisely because they are more efficient.

105. Three areas of regulation are relevant here.

106. First, the current regulation of vehicle standards includes standards intended to internalise externalities – in particular, by imposing ever-higher requirements in regard to emissions and fuel efficiency in new vehicles. By and large, this system has worked well, and particularly so in Europe, where it is has been applied most stringently. New vehicles are indeed increasingly cleaner and more fuel-efficient.

---

[35] Theoretically speaking, there is a trade-off that needs to be considered here. Regulating away low-quality products might entail pricing the product in question out of the reach of low-income households. Here, as elsewhere, regulation therefore needs to be subjected to a cost-benefit test. In practice, however, this particular regulation does not as yet present any major issues.

107. There is no reason to abandon such regulation following the introduction of comprehensive road pricing. Road-user charges differentiated by vehicle type will serve to reinforce the incentives that are already provided, by differentiated purchase taxes on vehicles – that is, they will help to speed up the conversion of the car fleet by encouraging motorists to switch from older, less clean and less fuel-efficient cars to newer, cleaner, more fuel-efficient replacements. Combining user-charges for motorists with emissions and fuel-efficiency standards for vehicle manufacturers would thus deliver a mutually reinforcing system of incentives for both consumers and producers. Hence, the recommendation in ECMT (2003) is to co-ordinate charging and regulatory systems in an optimal manner, not to replace one with the other.

108. Secondly, advanced traffic management systems are an essential complement to all road pricing schemes, they were essential to the success of the London congestion charge,[36] and will be essential to the success of future national schemes. In this regard, the introduction of pricing will entail an extension of regulation. But this should not occasion surprise. Advanced control systems are of course a permanent feature of scheduled services such as aviation and rail: demand is managed through a mix of regulation and pricing. Effectively managing demand in road use will likewise require a mix of regulation and pricing.

109. Thirdly, there is a need for a minimum of regulation of the pricing system itself. This applies to all modes, and it applies independently of the identity of the infrastructure operator, be it public authority, private firm or some form of hybrid public-private partnership. Once government has determined the new pricing rule and published an initial schedule of prices, there still remains the issue of determining actual prices on an on-going basis, as well as issues relating to the management of revenues. And decisions on these issues cannot be left to the discretion of the infrastructure operator, free of regulatory oversight.

110. But the minimum of regulation required to oversee the application of the pre-agreed pricing rule should not be confused with economic regulation in the conventional sense. There, the problem for the regulator is that of "optimising under constraint": the problem of determining, *in the absence of government intervention to facilitate welfare-optimal pricing,* how best to limit monopoly pricing power so as to minimise losses in welfare.[37] Here, the regulatory task is simpler: it is one of overseeing and enforcing the implementation of the welfare-optimal pricing that prior intervention has now made possible.

4. The constraints to optimising urban transport policy

4.1 The constraints to optimising pricing

111. The constraints to optimising urban transport policy, for each of its key constituent elements, can be grouped into three sets: economic and financial constraints; constraints arising from politics and civil society; and governmental, including especially inter-governmental, and related institutional constraints.

112. By far the most significant economic constraint to optimising pricing has been the absence to date of the technology required to do it.

113. In the case of the scheduled services of the railways, with its limited number of well-defined vehicle types and a more or less well-defined set of costs associated with each, it is indeed possible to price close to marginal social cost. Marginal internal costs can be tracked with relative ease. Marginal external costs are low, partly because environmental costs are low, but mainly because the external cost of

36 See TfL (2004) and the preceding and subsequent reports in the series.

37 For an exposition of the conventional regulatory problem and the limits to every available solution, see Armstrong, Cowan and Vickers (1994).

congestion is absent: the timetable prevents the excess train entering the system and imposing delays on others.

114. In the case of the road network, it is not possible to price close to marginal social cost without the advanced technology and related technical systems required to track and charge out the various internal and external costs imposed at the margin, including especially congestion. The costs imposed by each trip will vary widely depending on the route, the time of day and the type of vehicle. And the technology required so to differentiate the costs of each trip has hitherto not been available.

115. Thus, at the time of the renewal of the European policy debate on pricing, and judging that the requisite systems would not be available "in the foreseeable future", Verhoef (1999) was led to conclude that marginal social cost pricing was "nothing more than a hypothetical benchmark".

116. The conclusion does indeed follow from the premise. And if the option of marginal social cost pricing is simply not available because the marginal social costs in question cannot be tracked, it then becomes an open question whether an improvement upon the current situation is achievable by pricing rather than some other instrument. In this sense, the absence of the requisite technology becomes an economic constraint: economic criteria might well counsel against attempts to optimise prices relative to current settings.

117. To put the point another way: it is not at all clear that a general policy of second-best pricing would improve upon current outcomes. As Verhoef (1999) shows, such a route is far more complex and uncertain in outcome. And the modelled results for a range of European cities reported in Proost and Van Dender (2003) show not only that second-best options would deliver a significantly worse welfare outcome than would marginal social cost pricing, but that some would deliver a significantly worse outcome than what obtains today. Arguably, governments would be better advised to attempt incremental improvements with several instruments rather than pursue a general policy of second-best optimisation of prices.

118. But of course the premise is wrong. The satellite-based electronic systems required for implementing marginal social cost pricing are already available and operating in several European countries. They are now the favoured instrument used in kilometre-charging schemes for trucks on inter-urban routes. And the technical challenges required to adapt them for use in dense urban environments are not especially daunting. In the meantime, less sophisticated ground-based systems are widely in use across the world.

119. And once marginal social cost pricing becomes technically feasible, all outstanding economic and financial constraints – or economic and financial grounds for opposing a general policy of aiming at first-best optimisation – fall away in rapid succession.

120. As noted above, previous studies have shown that the expected overall gain in welfare is large. As reported in ECMT (2003), modelling for the year 2000 yielded results of as much as €13 billion (= £11 billion), €10 billion and €9 billion *per year* for Britain, France and Germany, respectively. And as external costs rise over time – in particular, the external costs of congestion and CO_2 emissions – so too do the gains from pricing these externalities. Thus, as reported in Eddington (2006), modelling for the UK for the year 2025, with a carbon cost of £95 per tonne, shows a welfare gain of £28 billion *per year*.

121. When viewed in this light, any doubts on the economic case for optimal pricing on the grounds of the high set-up costs of a sophisticated national road pricing scheme begin to vanish. As Eddington (2006) rightly remarks, "costs would have to be extremely high to outweigh the benefits of £28 billion a year."

122. Of course, this does not in itself dispense with the financial constraint: the set-up costs must be met.

123. Now the set-up costs of road pricing are an investment. And the general problem of how to overcome financial constraints to optimal investment is treated later. But since the financial constraint to this particular investment is also a constraint to optimal pricing, it is best addressed here as a separate issue.

124. Providing that prices are set at or close to marginal social cost, national road pricing will yield not only a large gain in welfare, but also a large gain in public revenue. For example, the first major national study by the UK Department of Transport, DfT (2004b), reported a gain of £9 billion per year compared to a welfare gain of £10-12 billion per year – which closely matched the equivalent estimates for the UK reported in ECMT (2003). [38] Hence, although it would be inadvisable for governments to blur the line between current and capital expenditure, Finance Ministers can rest assured on this one point: whatever the financial constraints to optimal investment in general, this particular investment is strongly financially-positive.

125. The evidence on revenues also helps to dispose of one of the earliest economic argument against marginal social cost pricing and the supposed economic constraint to marginal social cost pricing in rail: namely, that the subsidy it mandates in order to suppress a welfare loss for rail users can only be funded by raising welfare-reducing taxes elsewhere in the economy.[39]

126. The argument itself was always unsound. As noted earlier, pricing rail services without recourse to subsidy would entail over-charging rail users by up to 67%. There is scant evidence to suggest that the taxes required to fund this subsidy would impose a greater welfare loss than the welfare loss entailed by such a level of over-charging. Nor is it true that taxes are always welfare-reducing – taxes on the economic rents from land and natural resources do not reduce welfare.[40] Nonetheless, so long as the various markets of the transport sector were treated in isolation, the argument sometimes proved persuasive.

127. In the context of the current proposal to institute marginal social cost pricing in all modes of transport, the argument becomes irrelevant. For the transport sector as a whole, marginal social costs are higher than average producer costs. Hence, if all users are charged correctly, the transport sector as a whole delivers a revenue surplus. The subsidy required to prevent over-charging at some points in the system is thus paid for not by taxes from elsewhere in the economy, but simply by charging users correctly at every point in the same system.

128. If the economic and financial constraints to optimal pricing are destined to fall as the technology required to do it becomes available, the constraints arising from politics and civil society – in particular, the perception that public opinion is intractably opposed to the idea and that politicians would be well-advised to leave well alone – should have disappeared long ago. This perception is indeed a mind-forg'd manacle.

---

38 See Roy (2007) for a comparison of these two sets of estimates as well as others.

39 For a more recent revival of the debate on this point – but preceding the assembling of the new evidence base on revenue results – see *inter alia* Roy (1998), Rothengatter (2003) and Nash (2003).

40 Theoretically, once taxes on economic rents have been raised to their optimal level, the welfare-increasing effect of such subsidies would need to be compared with the welfare-reducing cost of taxes on capital, labour and consumption. But given that we are so far away from reaching this point in each and every OECD member-country, the particular counter-argument has little practical relevance.

129. The evidence contradicts the pessimism[41] – the evidence from polling, referenda and elections, and the experience of those cities and countries that have successfully implemented initiatives on pricing.

130. Perhaps the most comprehensive poll to date was that conducted by the UK's Royal Automobile Club Foundation[42] – its key relevant finding was that 71% of people find road pricing acceptable so long as it is introduced as part of a package of transport improvements. The first major European initiative, the Swiss Heavy Vehicle Fee, was carried by a 57% majority in the referendum prior to its introduction.[43] The most recent initiative, the Stockholm scheme, also passed its referendum test. London's Mayor Livingstone was elected in 2000 on a platform of introducing the Central London Congestion Charge and re-elected in 2004 following its introduction. And even when politicians are unable to carry the public on the issue, it is quite untrue to suppose that they must pay an electoral penalty: New York's Mayor Bloomberg keeps being re-elected despite his hitherto unsuccessful initiatives on road pricing.

131. Of course, as the example of New York shows, there is also evidence of public opposition to pricing schemes. Nor are all the polls which register opposition the result of inadequate methodology. But those who suggest that the settled view of public opinion will be negative falsely extrapolate from a particular moment in the play and project the result as if it were the end of the game. There is, as Goodwin (2006) describes it, a "gestation process for road pricing schemes". To begin with, there is no spontaneous public demand for pricing – rather there is an increasing concern with the concrete problems which can be solved only by pricing. Then, as the idea is explained as a solution to these problems, there is a build-up of support to the point where the idea is translated into a detailed proposal. At this point, there is a fall-off in support as various constituencies contest the detail. But "if there is sufficient steadiness of purpose to keep moving" and the scheme is brought to fruition, we witness a renewed build-up of support. And such steadiness of purpose is not without its reward. For the truth of it is that schemes that have been implemented are almost invariably popular.

132. A national road pricing scheme, as part of an optimal pricing policy for all transport modes, is a much larger undertaking. But the evidence above is equally relevant: public opinion is not a constant, let alone a constant that forbids the implementation of a policy that objectively delivers so much to the public.

133. The significant constraint to optimising pricing policy is not the public but government itself – more precisely, the division of powers and responsibilities between the various levels of government. And the levels of government that impact on urban transport are many: local, regional, national and international. For a local public authority seeking to optimise pricing policy, it is other governments that can often be the main constraint.

134. This is not to suggest that the constraint arises from any ill-will on the part of other governments. If it does, the matter of it lies outside the scope of the paper. What is relevant is the structure in which decision-makers are situated and assuming good-will on their part.

135. There are three main issues that need to be addressed – and need to be solved in practice.

136. First, local authorities must either possess or else acquire the powers required to carry through their responsibilities. For most local schemes, even if limited to city centres, this will require, at the least, the wherewithal to meet the set-up costs, and to maintain sufficient provision of public transport

---

41 The summary below borrows from the text of Roy (2007).

42 See RAC (2002)

43 See Balmer (2004).

alternatives so as to ensure that the pricing out of vehicles does not translate into a pricing out of people.[44] The fact that every scheme is, or ought to be, revenue-positive does not necessarily guarantee these results.

137. Set-up costs are, and should be treated as, an investment, and hence to be funded by borrowing and repaid from future revenues – they are not an item of current expenditure to be funded by current revenues. And local authorities may well lack the borrowing powers required. Hence, a higher level of government would need to play its part – either by meeting the set-up costs or else by extending local borrowing powers.

138. For cities with a well-developed public transport network, the power to ensure the sufficient provision of public transport alternatives is not principally a matter of funding. Revenues from pricing should provide the means to increase service frequency to the levels required to meet the additional demand.[45] But it does require decision-making powers in regard to public transport – for example, in regard to contracts with bus companies, service provision in the local segments of national rail networks, and so on. Not all cities will necessarily be in possession of these decision-making powers.[46] These will need to be acquired.

139. Secondly, local as opposed to national schemes, sometimes though not always, face a game-theory problem: whilst each would gain from the game if all were to play, any one may lose by playing if others refrain from doing so. The problem does not apply to the schemes in Central London or Stockholm or any other centre whose core economic functions cannot be readily duplicated elsewhere. But for comparable economic centres, especially those at close distance to one another, there is some risk of businesses relocating from the centre that initiates pricing to those that do not. One solution is to introduce local-area pricing in a sufficiently co-ordinated manner and with strong support from national government – a strategy that is now being pursued in the UK. The more comprehensive solution is of course to introduce a comprehensive national scheme – the strategy that has been adopted in the Netherlands.

140. More generally, it needs to be recognised that cities cannot, by local action alone, fully implement, and fully realise the benefits from, optimal pricing in urban transport. Local authorities are not, nor are likely to be, in a position to determine the full set of prices in the transport system within their own area of jurisdiction, including the prices on all local segments of national networks. But what cities can do is to form a powerful constituency in favour of implementing optimal pricing at a national level – and help to trigger that decision by implementing exemplary local pricing schemes to the best of their ability. After all, as was noted earlier with reference to the cross-country evidence from Europe, by far the greater part of the overall gain from optimal pricing will accrue to urban areas.[47]

141. Finally, even in the event of such a national decision, cities will continue to be impacted by the decisions of other nations – in particular, by the success or failure to reach international agreement on

---

44 Theoretically, there are other alternatives, such as car-sharing. In practice, the experience of both London and Stockholm shows it is public transport that absorbs the better part of the displaced demand. And car-sharing plays little part in the story. See Roy (1997) and the sources cited therein.

45 Of course, such cities may also need to invest in additional public transport capacity either to make up for past shortfalls or in anticipation of future needs or both. But that is a separate issue to the problem under discussion.

46 London is; most British cities are not.

47 The evidence is most striking in regard to the UK where 88% of the overall welfare gain accrues to urban areas – see above and ECMT (2003). Greater London alone is likely to capture more than 50% of the overall gain from national road pricing, dwarfing by far the gains from the Central London Congestion Charge.

action to combat climate change. It is, however, the national decision on pricing, including the pricing of carbon and the determination of its shadow price in investment appraisal, which is most critical for outcomes in urban transport.

4.2 The constraints to optimising investment

142. Strictly speaking, there is no *economic* constraint to – and no economic case against – the optimal investment rule specified above: proceed if, and only if, NPV > 0. To fail to invest in projects that offer positive net present values, calculated on the basis of accurate discounting, is not "prudence" – rather the contrary. To say that we "cannot afford" to invest, *after taking into account the value of the sacrifice of present consumption*, is simply a disguised way of admitting that we "cannot afford" to consume as we do.

143. Potentially, there are of course *financial* constraints to implementing such an optimal investment policy. The authority in question must be able to borrow in sufficient quantity and at an affordable price (interest rate). Moreover, if the chosen discount rate is less than the long-term cost of borrowing, it must be able to absorb the impact of the difference. And if these conditions are not met, the negative financial consequences will indeed have an economic impact.

144. Hence, in the name of "realism", many macro-economic experts have counselled, and many Finance Ministries have concurred, that such an investment policy is unaffordable and attempting to implement it would be irresponsible.

145. In fact, the evidence from the real world tells another story. In the course of the last three decades, the economic take-off in China and India and the rest of industrialising Asia has vastly expanded the world's pool of savings and driven long-term interest rates to historical lows. The cost of borrowing for OECD governments has never been lower. The yields on long-dated gilts issued by the UK's Treasury with 30-50 year maturities fell to less than 1% in early 2005. In September 2008, yields on some US Treasury bills turned negative!

146. Now the scale of public investment, though large in relation to the sectors in which it is required, is small in relation to the macro-economy. For all practical purposes, throughout this period, there was no financial constraint in operation. This too was another mind-forg'd manacle.

147. The consequence of OECD governments failing to borrow in sufficient quantity to fund worthwhile investments in infrastructure and other transport improvements, as well as other worthwhile investments in all the fields for which government is properly responsible, is now becoming clear. The excess savings of the emerging economies were translated not into infrastructural and other worthwhile assets but current expenditure by government and financial "assets" that fed a housing bubble in the advanced economies and then an inevitable housing bust. And after the bust, savings from the same source will now be used by the US and other OECD governments to buy out these financial "assets" and detoxify the financial system.

148. Looking ahead, and beyond the duration of the current financial crisis, the question at issue is how the US will use the share of world savings that flows into US Treasury bills. The welfare criterion suggests that the small fraction of it required to fund an optimal investment policy is indeed used for this purpose.

149. In the meantime, in the US and across the OECD world, the utility of public investment, in transport and elsewhere, could be greatly increased by adopting a discounting regime that accurately recognises long-term costs and benefits and prioritises projects accordingly. If we cannot proceed with all worthwhile investments, we can at least ensure that the most worthwhile ones are placed at the front of the queue.

150. Optimising investment policy, and in the first place optimising the priority list, will continue to face constraints arising from politics and civil society, particularly in the context of planning applications for major projects. But these are highly localised and relatively weak constraints. They can be overcome by thorough preparation, clear communication and strong leadership

151. In a larger context, the ideological climacteric of 2008 should prove immensely helpful to the task of moving toward an optimal investment policy. Public opinion, in the US and elsewhere, is likely to place a higher value on public investment – in transport and other infrastructure, including social infrastructure – and a lesser value on private investment in certain classes of assets.

152. What will remain to be negotiated are the constraints arising from inter-governmental relations. The argument here is essentially the same as spelt out above and does not need to be rehearsed again. In regard to investment, as in regard to pricing, powers and responsibilities must be matched. And this is indeed one of the key challenges that Australia is now addressing at the outset of its multi-billion dollar programme to renew the nation's infrastructure across all sectors.

4.3 The constraints to optimising regulation

153. As noted earlier, there are three areas of regulation that are most relevant to the problem at hand: the tightening of vehicle standards, the extension of traffic management, and the regulation of the new pricing system. In each of these, the pursuit of an optimal policy will face some measure of constraint; in none of these are the constraints insurmountable.

154. The tightening of vehicle standards is a programme that has been successfully implemented over the years in European countries. So long as the appropriate cost-benefit tests are met, the economic case stands. There are no financial constraints for government. And as yet there are no unintended macro-economic consequences to note: vehicle manufacturers have not collapsed under the burden

155. Civil society has been largely supportive.

156. There is an inter-governmental issue here, but it is not one between the levels of government in Europe. Rather, it is the fact that there is a wide divergence of standards between the several various vehicle-manufacturing countries, and this impacts on local roads under conditions of free international trade. Hence, it would be helpful to move toward a common set of standards over time – that is, by others moving up toward the European level rather than Europe moving down toward the level of others.

157. The constraints to the extension of traffic management, including the introduction of new, more advanced, systems of management, are weak. Of course, the programmes must be funded. And local authorities will need to ensure that any financial constraints are negotiated away. Public education ought to be able to deal with any public disquiet. And the only inter-governmental issues likely to arise are to be found within the components of local government itself.

158. The economic and financial constraints that apply to the conventional regulatory task should not apply to the regulation of the new pricing system. It is no longer a question of how to minimise welfare losses. The task here is to ensure the implementation of the welfare-optimal rule. Whilst there are political challenges to overcome in order to establish the new system, the system itself is rule-bound and de-politicised. There are no obvious inter-governmental issues.

159. But it would be as well to enter at least two caveats. First, since what is being proposed here is a new system, re-assurance cannot be found in past experience. New regulatory challenges might well emerge of which we have no experience.

160. Secondly, and more specifically, there is the issue of existing institutions and contracts in road, rail and other public transport – including the issue of long-term contracts with private-sector operators that stretch well beyond current proposals to introduce national road pricing and parallel pricing rules in all modes of transport. In principle, this should not delay or disrupt the institution of welfare-optimal pricing – even if there is a financial impact to be absorbed in any necessary adjustment to existing contracts. Indeed, the fact of such contracts should be viewed in a positive light – that is, as a legally-binding mechanism for re-enforcing government's obligations to pay the subsidies required to permit the operation of the P = MC rule. But once again new issues could emerge as the process unfolds.[48]

161. That said, it remains to repeat that none of the current and potential constraints to optimal regulation appear insurmountable.

5. Conclusion

162. This paper, like the process of policy reform in Europe which it has sought to articulate, has been a long journey from theory to policy and thence to some of the key challenges facing those who are charged with delivering the results on the ground.

163. But its message to the public authorities responsible for urban transport is unambiguous and unambiguously positive.

164. Relative to what obtains today, an optimal policy on transport pricing, investment and regulation will deliver very large gains. It is perfectly possible to overcome the constraints to optimising policy. And it is time to call some of the supposed constraints for what they are – imagined, mind-forg'd manacles from which we can break free.

[48] For some indications of potential pitfalls, see OECD/ITF (2007). Nonetheless, the report insists, and rightly so, that the efficiency (*i.e.*, welfare-optimality) criterion should predominate.

REFERENCES

ARE (2004), *Fair and Efficient: The Distance-related Heavy Vehicle Fee (HVF) in Switzerland*, Berne.

Armstrong, M., S. Cowan and J. Vickers (1994), *Regulatory Reform: Economic Analysis and British Experience*, The MIT Press, Cambridge, Mass.

Arrow, K. and G. Debreu (1954), "Existence of an Equilibrium for a Competitive Economy", *Econometrica*, Vol. 22, No. 3, pp. 265-290.

Balmer, U. (2004), The window of opportunity: How the obstacles to the introduction of the Swiss heavy goods vehicle fee (HVF) have been overcome, prepared for the OECD, Paris. Available at http://appli1.oecd.org/olis/2004doc.nsf/linkto/com-env-epoc-ctpa-cfa(2004)57-final.

Brittan, S. (2008), "Capitalism and the credit crunch", *The Financial Times*, 11 September 2008, London.

Caldwell, C. (2008), "There's no free lunch and no free economy", *The Financial Times*, 19 September 2008, London.

Caves, D. N., L.R. Christensen, M.W. Tretheway and R. J. Windle (1985), "Network Effects and the Measurement of Returns to Scale and Density for US Railroads", in A. Daughety, ed., *Analytical Studies in Transport Economics*, Cambridge University Press, Cambridge.

CE Delft (2002), *Returns on Roads: Optimising road investments and use with the 'user-pays principle',* Delft.

DfT (2004a), *The Future of Transport: a network for 2030*, Department for Transport, London.

DfT (2004b), *Feasibility Study of Road Pricing in the UK: A report to the Secretary of State for Transport*, Department for Transport, London.

DfT (2006), *Transport, Wider Economic Benefits and Impacts on GDP*, Department for Transport, www.dft.gov.uk.

DfT (2007a), *Delivering a Sustainable Railway*, Department for Transport, London.

DfT (2007b), *Delivering a Sustainable Railway: Summary of Key Research and Analysis*, Department for Transport, www.dft.gov.uk.

DfT (2007c), *Toward a Sustainable Transport System: Supporting Economic Growth in a Low Carbon World*, Department for Transport, London.

EC (1995), *Towards fair and efficient pricing in transport: Policy options for internalizing the external costs of transport in the European Union*, Green Paper, Commission of the European Communities, Brussels.

EC (1998), *Fair payment for infrastructure use*, White Paper, Commission of the European Communities, Brussels.

EC (2006), *European Transport for 2010*, White Paper, Commission of of the European Communities, Brussels.

EC (2008a), *Communication from the Commission: Strategy for the internalisation of external costs*, Commission of the European Communities, Brussels.

EC (2008b), *Commission Staff Working Document: Summary of the Impact assessment on the internalisation of external costs*, Commission of the European Communities, Brussels.

ECMT (2001), *Assessing the Benefits of Transport*, OECD Publications, Paris.

ECMT (2003), *Reforming Transport Taxes*, OECD Publications, Paris.

Eddington, R. (2006), *The Eddington Transport Study, Main report: Transport's role in sustaining the UK's productivity and competitiveness*, HM Treasury, London.

FT Editorial Comment, "Changing the rules of the game", *The Financial Times*, 17 September 1008, London.

Glaister, S. (1994), "Public transport: The allocation of urban public transport subsidy", in R. Layard and S. Glaister, (eds.), *Cost-Benefit Analysis*, Cambridge University Press, Cambridge.

Goodwin, P. (2006), "The gestation process for road pricing schemes", *Local Transport Today*, LTT444.

HMT (2003), *The Green Book: Appraisal and Evaluation in Central Government*, Treasury Guidance, HM Treasury, London.

Hotelling, H. (1938), "The General Welfare in Relation to Problems of Taxation and of Railway and Utility Rates", *Econometrica*, Vol. 6, No. 3, pp. 242-269.

Hotelling, H. (1939), "The Relation of Prices to Marginal Costs in an Optimum System", Econometrica, Vol. 7, No. 2, pp. 151-55.

Kessides, I. N. and R. D. Willig (1995), *Restructuring Regulation of the Rail Industry for the Public Interest*, Policy Research Working Paper 1506, The World Bank, Washington, D.C.

Lange, O. R. (1942), "The Foundations of Welfare Economics", *Econometrica*, Vol. 10, No. 3/4, pp. 215-228.

Layard, R and S. Glaister, (Eds.) (1994), *Cost-Benefit Analysis*, Cambridge University Press, Cambridge.

Lerner, A. P. (1934), "The Concept of Monopoly and the Measurement of Monopoly Power", *The Review of Economic Studies*, Vol. 1, No. 3, pp. 157-175.

Lerner, A P. (1944), *The Economics of Control: Principles of Welfare Economics*, McMillan (1960), London.

Lipsey, R. G. and Lancaster, K. (1956 - 1957), "The General Theory of Second Best", *The Review of Economic Studies*, Vol. 24, No. 1, pp. 11-32.

Mohring, H (1972), "Optimisation and Scale Economies in Urban Bus Transportation", *American Economic Review*, Papers and Proceedings, 62, pp. 591-604.

Nash, C. (2003), "Marginal cost and other principles for user charging in transport: a comment", *Transport Policy* 10, pp. 345-348, www.elsevier.com.

OECD (2005), *Environmentally Harmful Subsidies: Challenges for Reform*, OECD Publications, Paris

OECD/ITF (2007), *Transport Infrastructure Investment, Options for Efficiency*, OECD/ITF, Paris.

OECD/ITF (2008), *The Wider Economic Benefits of Transport: Macro-, Meso and Micro Transport Planning and Investment Tools*, OECD/ITF, Paris.

Pigou, A. C. (1912), *Wealth and Welfare*, McMillan (1920), London.

Proost, S. and K. Van Dender (2003), *Marginal Social Cost Pricing for all Transport Modes and the Effects of Modal Budget Constraints*, ETE Working Paper Series, no. 2003-11, Leuven.

RAC Foundation (2002), *Motoring Toward 2050: an independent inquiry*, London.

Rothengatter, W. (2003), "How good is first best? Marginal cost and other pricing principles for user charging in transport", *Transport Policy* 10, pp. 121-130, www.elsevier.com.

Roy, R. (1998), *Infrastructure Cost Recovery under Allocatively Efficient Pricing*, UIC, Paris.

Roy, R. (2001), "Means and Ends: Cost-Benefit Assessment and Welfare-Maximising Investment", in ECMT, *Assessing the Benefits of Transport*, OECD Publications, Paris.

Roy, R. (2005), *The impact of pricing reform on investment needs: re-writing the reference scenario*, prepared for the ECMT, Paris.

Roy, R. (2006a), "The public-private partnership and the general welfare", in *The Railways, the Market and the Government*, Institute of Economic Affairs, London.

Roy, R. (2006b), *Investing in the new century: Toward an undistorted appraisal process*, Report for the Railway Forum, London.

Roy, R. (2007), *A Politician's Guide to Efficient Pricing*, ECMT, Paris.

SACTRA (1999), *Transport and the Economy*, TSO, London.

Samuelson, P. A. and W. D. Nordhaus (1995), *Economics*, Fifteenth Edition, Irwin/McGraw-Hill, Boston, Mass.

Sansom, T., C. Nash, P. Mackie and J. Shires (2001), *Surface Transport Costs and Charges: Great Britain 1998*, Final Report for the Department of Environment, Transport and the Regions, ITS, Leeds.

Schumpeter, J. A. (1911), *An Inquiry into Profits, Capital, Credit, Interest and the Business Cycle*, Harvard University Press (1934), Cambridge, Mass.

Schumpeter, J. A. (1942), *Capitalism, Socialism and Democracy*, Unwin (1987), London.

Spackman, M. (2001), "Public Investment and Discounting in European Union Member States", *OECD Journal on Budgeting*, Vol. 1, No. 2, pp. 213-260.

Spackman, M. (2004), "Time Discounting and of the Cost of Capital in Government", *Fiscal Studies*, Vol. 25, No. 4, pp. 467-518.

Stern, N. H. (2006), *The Economics of Climate Change: The Stern Review*, Cambridge University Press (2007), Cambridge.

Sterner, T. and U. M. Persson (2007), *An Even Sterner Review: Introducing Relative Prices into the Discounting Debate*, Discussion Paper, Resources for the Future, Washington, DC.

Stockholm Stad (2006), *The Stockholm Trial*, Presentation to ECMT Conference on Road Charging Systems, 1 June 2006, Paris, France, www.cemt.org/topics/taxes/Paris06/.

TfL (2004), *Central London Congestion Charging: Second Annual Report*, Transport for London, London.

TfL (2005), *Central London Congestion Charging: Third Annual Report*, Transport for London, London.

TfL (2006), *Central London Congestion Charging: Fourth Annual Report*, Transport for London, London.

Verhoef, E. T. (1999), "The implementation of marginal external cost pricing in road transport", *Papers in Regional Science*, Vol. 79, pp. 307-332.

Vickerman, R. and J. Monnet (2001), "Transport and Economic Growth", in ECMT (2001), *Assessing the Benefits of Transport*, OECD Publications, Paris.

Walras, Léon (1877), *Elements of Pure Economics*, Harvard University Press (1954), Cambridge, Mass.

Werther, P. (2006), *KM pricing in the Netherlands*, Presentation to ECMT Conference on Road Charging Systems, 1 June 2006, Paris, France, www.cemt.org/topics/taxes/Paris06/.

Zupan, J. and A. Perrotta (2003), *An Exploration of Motor Vehicle Congestion Pricing in New York*, Regional Plan Association, New York, NY.