UNITED
NATIONS

E

Economic and Social
Council

Distr.
GENERAL

ECE/TRANS/2007/7/Add.1
5 January 2007

ENGLISH
Original: ENGLISH and FRENCH

ECONOMIC COMMISSION FOR EUROPE

INLAND TRANSPORT COMMITTEE

Sixty-ninth session
Geneva, 6 - 8 February 2007
Item 10(d) of the provisional agenda

TRANSPORT TRENDS AND ECONOMICS

Studies on transport economics and track costs undertaken by other organizations

Round Tables 135, 136 and 167

Submitted by the Organisation for Economic Co-operation and Development (OECD) and the European Conference of Ministers of Transport (ECMT) [*/]

Appendix 1: Conclusions of Round Table 135:
Transport Infrastructure Charges and Capacity Choice;

Appendix 2: Conclusions of Round Table 136:
Estimation And Evaluation of Transport Costs;

Appendix 3: Conclusions of Round Table 137:
Transport, Urban Form and Economic Growth.

[*/] The information in this document is reproduced in the form in which it was received by the Secretariat.

APPENDIX 1

CONCLUSIONS OF ROUND TABLE 135: TRANSPORT INFRASTRUCTURE CHARGES AND CAPACITY CHOICE

INTRODUCTION

A Round Table on "Transport Infrastructure Charges and Capacity Choice" was held on 29-30 September 2005 at the OECD in Paris. The meeting was chaired by Werner Rothengatter (University of Karlsruhe and President of the World Conference of Transport Research Society). Background papers were provided by Georgina Santos (Oxford University), Erik Verhoef (Free University of Amsterdam and Tinbergen Institute), Ken Gwilliam (Department of Transportation, Florida) and Barry Potter (Director at the International Monetary Fund).

The White Paper of the European Commission. (EC, 2001) initiated a broad and extended discussion was in Europe on charges for infrastructure use. A similar, but narrower discussion continues in the USA on "value pricing" of road use, i.e. extending the existing road system by the provision of fast lanes which will be charged for, while other lanes can be used for free (Small, 2001). These discussions had focused on charges for infrastructure use as a means to contain transport demand, and on the internalisation of external costs, mainly environmental and safety, by relating the level of charges to these costs at the margin. Despite extensive discussion, many stakeholders and policy analysts are disappointed by the level of implementation. Where infrastructure charges have been introduced -- the London toll ring and the German motorway charges for heavy duty trucks -- the pricing rules employed differ markedly from the concepts recommended by the policy planning concepts (Prud'homme and Bocarejo, 2005). The level of political resistance to a more general introduction remains high. In many cases, fiscal motivations for introducing charging schemes were as important as objectives to increase the efficiency of the transport sector.

The fiscal reasoning for the introduction of infrastructure charging is based on the claim that only revenues from infrastructure charging could ensure that sufficient resources are available to finance the transport infrastructure required to achieve economic development objectives. This claim is, sometimes implicitly, associated with the argument that these resources cannot be generated from general taxes. This, in turn, is due either to the impossibility to increase taxes or the fact that transport policy allocations fall short of what an optimal fiscal plan would indicate.

Against the backdrop of this situation, the Round Table addressed the following questions, which have been neglected in much of the earlier discussions:

-- Is it possible to design a quasi-market of infrastructure services? Such a market would have to be based on prices for infrastructure services which are not only aimed at target demand levels, but which guide supply decisions at the same time. The supply of infrastructure services would be determined by the decision on the capacity of the infrastructure stock.

-- Would the pricing rules that have been discussed in recent debates have to be modified? This would mainly concern the question of to what extent external costs would and should be taken account of in defining a pricing rule that would help to provide infrastructure services at the least cost possible.

-- Would charging and the provision of infrastructure services at least cost be possible by relying exclusively on distance-related charges? If not, how should the difference between full costs and the prices required for optimal use be covered?

-- Is political resistance to road pricing inevitable? The introduction of infrastructure charging has been argued for by presenting it as an efficiency enhancing step of fiscal reform. Infrastructure users perceive the introduction of infrastructure charges as just another tax. The gap between the implementation of charging systems and its justification -- e.g. pricing according to broad vehicle classes which are only loosely connected to environmental costs – has led to the perception by at least some user groups, that infrastructure “pricing” would in fact mean the introduction of just another tax with a justification lacking plausibility and credibility.

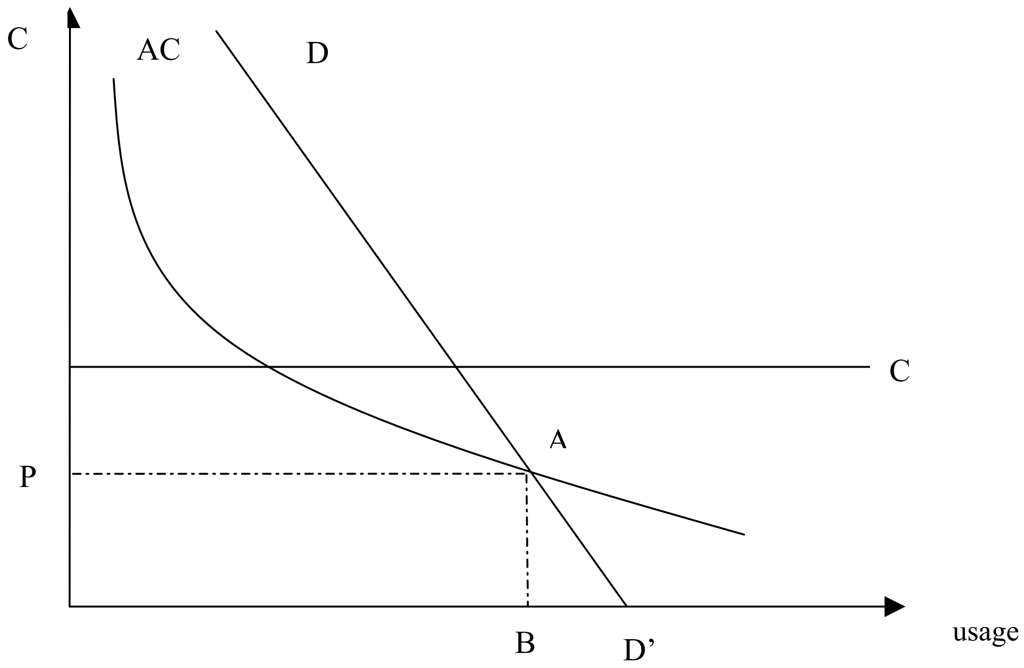
-- What mechanisms are required to ensure that the revenues of infrastructure charging are indeed used as payment for infrastructure services and not diverted to other purposes? This question is closely related to the debate on the desirability of infrastructure funds, most prominently the one between the World Bank and the International Monetary Fund. That discussion focused on how to ensure an appropriate funding of road maintenance. While the World Bank sought a remedy to cure the underfunding of road maintenance in some client countries, the IMF was concerned that off-budget cost recovery mechanisms would increasingly interfere with rational fiscal programming. There were also doubts that the abuse of infrastructure funds for other purposes could be avoided. The Round Table discussion showed that beliefs about the likely outcomes of budgetary processes did not differ as much as earlier discussion had suggested and that there was agreement on the need for careful institutional design of the management of road funds to respond to the concerns of the IMF.

PRICING PRINCIPLES AND COST RECOVERY

It might be instructive to recall why transport infrastructure services are normally not provided on private markets, and why local or national governments play a major role in taking infrastructure investment decisions to build up capital stock which determine the capacity of individual facilities and networks overall. The salient feature of transport infrastructure services is the fact that they require a durable input good -- transport infrastructure capital stock -- which can only be installed in large, discrete units. Once established, the investment is sunk, i.e. there is no possibility to turn infrastructure capital stock into another physical good, nor does a resale market normally exist for these capital goods. These characteristics imply that, at least at low levels of demand, the costs per user of the infrastructure decrease with an increase in the number of users and the level of service demanded by the individual user.

What this means for pricing may be clarified by an extreme example. It is possible that, in sparsely populated geographical areas, all of the infrastructure costs are independent of the actual low levels of usage. The initial construction costs would imply a certain annual opportunity cost of finance (debt service or other public service forgone by having built the infrastructure facility) and administrative as well as maintenance costs are determined by the fixed annual costs of a minimum administrative and maintenance unit. Different use levels of, let's say, a road system by the local population do not make a difference to the routine activities of the administration and maintenance unit. In this case, user charges per kilometre of road use would not improve the provision of infrastructure services but worsen it, as may be illustrated by Figure 1.

Figure 1. Only fixed costs: optimal prices are zero



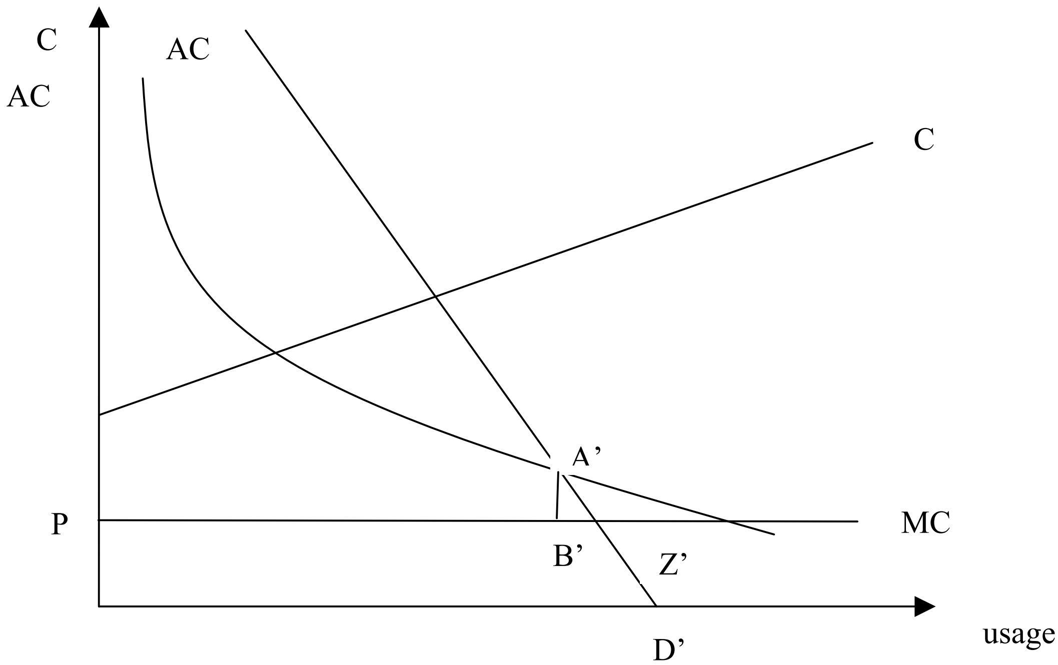
The average cost curve (AC) indicates the decreasing costs which accrue annually per unit of service, depending on the total demand for services. As all annual costs are fixed (C), they decrease continuously with an increasing use level. DD' indicates demand for infrastructure services depending on charges per unit of usage. The area below the demand curve indicates the benefits resulting from the use of the infrastructure facility. Any price charged would lead to a loss of benefits, due to the reduction in road use that is associated with a price of zero, without leading to a reduction in costs (being all fixed due to low levels of usage). Should a provider be required to provide the service at prices per unit of service that just cover costs, the price would be P. At this price level, the users would lose benefits equivalent to the triangle (ABD') without any associated reduction in costs. The optimal per-km user charge in such a case would be zero. While this might explain why there are normally no private markets to provide infrastructure services, the above argument does not imply that a provider cannot or should not recover the (fixed) costs of providing the service. The optimal way of recovering costs is to levy a lump-sum access charge for all users of the transport infrastructure facility, i.e. a charge that is unrelated to individual usage[1/]. The decision on capacity choice would be a trivial one. The road administrators would invest in a facility of minimum size (e.g. two lanes) and a desirable quality.

1/ If there are major differences between users in vehicle types used, a distributional problem can arise. On how to respond to this problem by differentiating fixed charges cf. Kopp (forthcoming).

A more important reason for government involvement in the provision of infrastructure is the fact that the unregulated private provision of infrastructure would lead to an underutilization of the resources that have been invested in transport infrastructure stock. In fact, even if all infrastructure costs were fixed costs, an unregulated provider would increase the price up to a point where the effects of the price increase on revenues and the induced decrease of demand cancel out. The price would be much higher than P in Figure 1. In other words, a (monopolistic) provider will seek to turn consumer benefits into rents and thereby reduce the benefits to society overall[2/].

This argument remains valid if the administration and maintenance costs increase with an increase in the number of users or an increase in infrastructure use per user, if the increases are constant. A linear relationship between infrastructure costs and usage, i.e. constant marginal costs, still leads to decreasing average costs due to the fixed construction and maintenance costs, as shown in Figure 2. The slope of rays from the origin of the co-ordinate system to the cost function, indicating the average costs, decreases with an increase in usage.

Figure 2. Constant marginal costs: optimal prices don't cover full costs



With positive marginal costs, additional use of the facility will directly lead to additional costs. If users are not charged for these additional costs, they will tend to overuse the facility. If a cost-recovery mechanism like a lump-sum access charge were in place, users with very high individual demand would cause additional, use-dependent costs that could partly be shared by users with low demand, above the level to which they have contributed to variable costs. To avoid incentives to overuse the infrastructure in this sense, and to avoid the negative distributional consequences, the price per unit of infrastructure use should be equivalent to the (constant) marginal costs.

[2/] Depending on the provider's opportunities to price discriminate, different prices for different users could be implemented to turn consumer benefit into profits of the providing firm.

The revenue from pricing would still not cover the full costs: the fact that the average costs decrease indicates that the additional costs following from additional usage fall short of the unit costs. A price higher than the marginal costs would, however, indicate an under-utilisation of the accumulated stock of (sunk) investment. The collective of users would lose benefits equivalent to the triangle A'B'Z'. To cover the total costs, a fixed access charge, for example a vehicle tax, is required that covers the difference between total costs and the revenues from marginal cost pricing. As long as congestion continues to be unimportant, a linear cost function for infrastructure use does not alter the above argument on capacity choice: infrastructure managers will choose the minimum capacity, as expected demand remains below the level where time losses, due to crowding, occur. As in the case with only fixed costs, an unregulated provider will try to implement charging policies that do not only cover fixed costs but that maximize monopoly rents.

EXTERNAL COSTS OF INFRASTRUCTURE USE

External road damage costs

As mentioned in the introduction, external costs, i.e. costs generated by one group of infrastructure users that accrue to other users or to non-users, and their inclusion in the calculation of prices, have played a major role in pricing discussions. The environmental costs of most political concern are those that affect non-users. The external part of road damage and congestion costs concern other road users. As we will see, pricing schemes to recover costs should include some of these costs, but should exclude others if the prices and revenues in equilibrium are supposed to give guidance to investment policies.

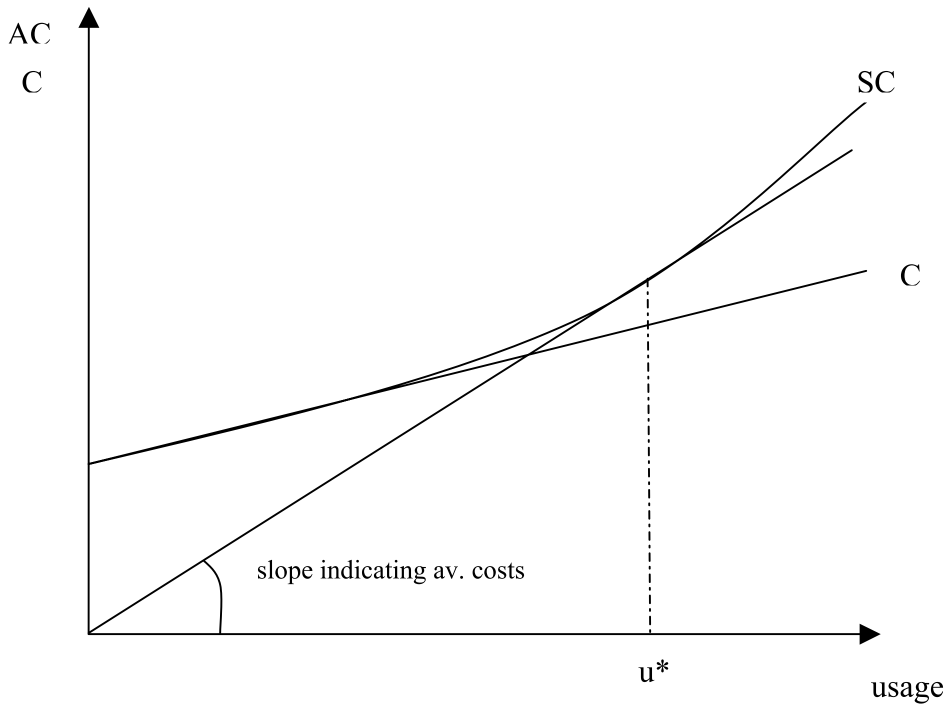
A first category of "external" cost, closest to the core administrative and maintenance costs, arises from road damage (Newbery 1988). Similar externalities may hold for other types of transport infrastructure. When a vehicle damages the road surface, the increased roughness of the surface leads to increased *vehicle costs* for subsequent traffic. The increased vehicle operating costs constitute a road damage externality. On well-trafficked inter-urban roads these vehicle operating costs are estimated to be between 10 and 100 times as high as maintenance costs (Newbery, 1988, p. 298). This has not generally been considered in debates on the internalisation of external costs in road pricing to date.

Road damage caused by vehicles advances the date at which the road needs to be repaired. The most important type of damage is best expressed by the increased roughness of the surface, which can be quantified by instruments like the Bump Integrator. For a well-designed road, its initial roughness will be low and will continuously increase with the passage of traffic. Depending on pavement type the roughness reaches, after 10-20 years, a level at which major maintenance, such as an asphalt overlay, is required to restore the surface to its initial low level of roughness.

The damage vehicles cause depends on the type of vehicle and the type of road. The damage of vehicle types can be measured as some fraction or multiple of a standard damaging unit as, for example, the equivalent standard axle load for paved roads.

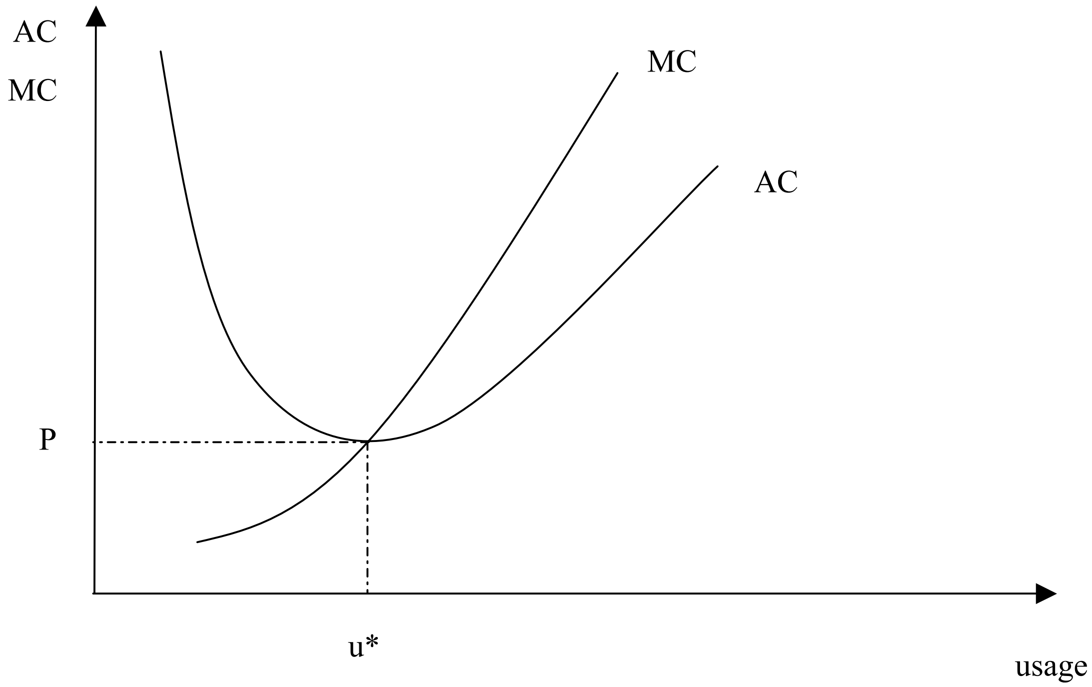
What is important for the discussion of the link between pricing and investment decisions is the fact that adding the external road damage costs to the cost function of road infrastructure services makes the total cost function convex to the origin. In other words, the additional (internal and external) costs of transport infrastructure services increase with increasing demand. This has the important consequence that the total infrastructure costs per user or per unit of service no longer -- as in the cases of only fixed costs or fixed costs and constant marginal costs -- decrease continuously with increasing demand. Due to the fact that the increase in marginal costs is weak for low levels of traffic, an increase implies that the average costs might decrease for low levels of demand, reach a minimum and increase with strongly increasing external road damage costs, as shown in Figure 3.

Figure 3. Infrastructure Costs and Road Damage Externality



That is, even if the cost function, which comprises only administrative and maintenance costs, should be linear, the external road damage costs in the form of increased vehicle operation costs, for example, due to an increased roughness of road surfaces, might lead to an exponentially, i.e. more than proportionately, increasing total cost function. The minimum of the associated average cost function is associated with the equality of average and additional costs caused by the last extra unit of service, indicated by u* in Figure 3. Figure 4 shows the marginal cost curve, MC, indicating additional maintenance and administrative costs as well as vehicle damage, resulting from increased roughness of road surfaces due to an additional unit of road service, and the average cost curve AC, indicating total costs per unit of service. It illustrates that, for low levels of usage, average costs are higher than marginal costs. That is, for levels of usage below u* its increase reduces the amount of resources required per unit of service. In other words, the efficiency of the infrastructure sector is increased as the resources bound by past investment and spent for maintenance allow an increase in transport services.

Figure 4. Average Costs and Optimal Capacity



Pricing according to marginal cost at u* would just cover the average costs. At this usage level, charging only for the additional internal and external costs covers the full costs of the facility, despite the fact that transport infrastructure facilities are usually associated with relatively high, fixed construction costs. For demand that is at least as large as the usage at minimum average costs (u*), per-unit-of-service charges cover the full costs.

That demand at a minimum average cost price exactly happens to be equal to the minimum cost level can only happen by chance.

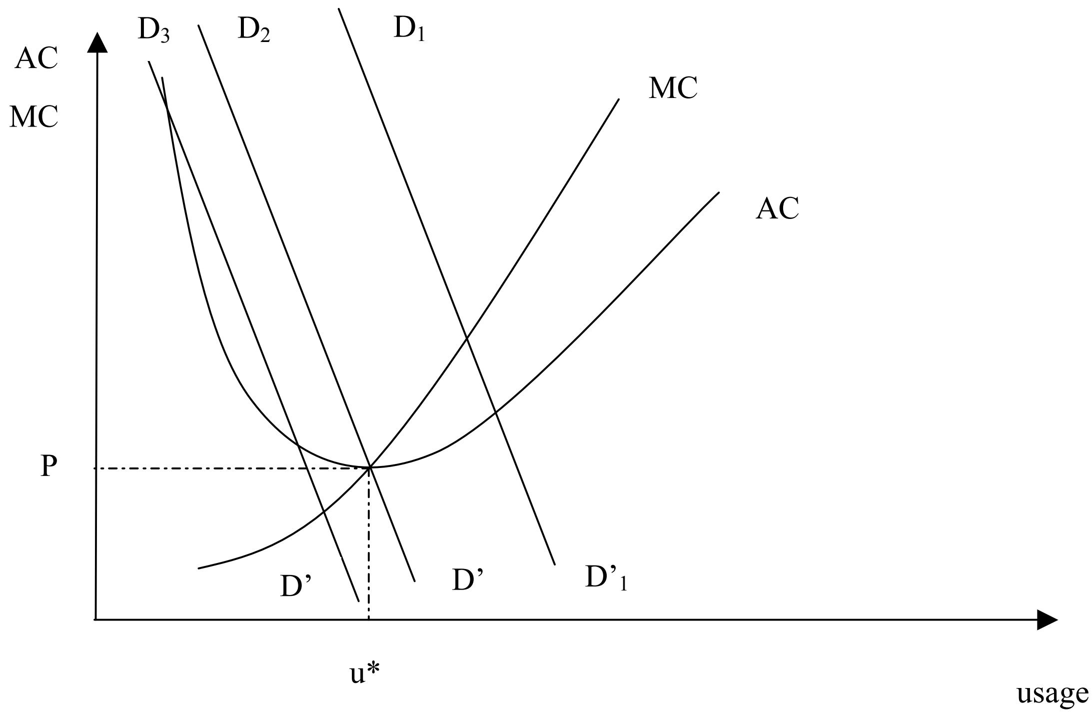
For the decision on the capacity of the facility, this can have three possible consequences:

1. The demand level is smaller than the minimum cost use level. In that case, the existence of a minimum average cost is irrelevant for infrastructure managers. An increase in usage will lead to a decrease in average costs, and therefore the above arguments for the cases of only fixed costs and proportional additional costs apply. The price should not cover the full costs, as it would lead to an underutilisation of the infrastructure facility. A fixed charge (e.g. an access charge, a vignette or a vehicle tax) should be used to cover the difference between total costs and revenues from minimum AC. Knowing that demand is and will remain low, the pricing infrastructure managers will choose the minimum capacity.

2. Demand just equals the use level. A price which just covers the additional costs resulting from an extra unit of demand will cover the full costs. The capacity of the facility equals the minimum average costs.

3. Demand at the minimum average cost price is greater than the optimal use level. In that case, a marginal cost price will lead to revenues which are greater than the total costs of the facility. At the same time, infrastructure users will suffer from high levels of vehicle damage due to the high level of wear and tear of infrastructure facilities. With increasing usage, it will become cheaper for society overall to add lanes or establish a second facility leading to lower costs per service unit, despite the need to finance a second block of fixed costs, because it avoids the high external costs of vehicle damage.3 With an increase in the number of facilities (road lanes, railway tracks), the demand per facility might fall short of optimal use levels u*. The per-unit-of-service price should always be equivalent to minimum average costs, and possible deficits covered by a fixed fee.

Figure 5. Optimal Capacity and Equilibrium Demand



Congestion Costs

Congestion costs have been more important than the vehicle costs resulting from overuse of infrastructure for the development of pricing rules as well as for practical policy discussions. The quantity dimension of congestion costs is measured in time units as the delay caused by a number of simultaneous users of a facility which impedes a free flow of vehicles (defined by adequate safety standards). A long and protracted discussion exists on how to value these delays (cf. Round Table Report 127). The Round Table discussed the role of congestion costs for rules of infrastructure pricing and decisions on the capacity of infrastructure facilities.

3/ The fact that without charges, and without a proper accounting of external costs the perceived demand by planners has often been too high relative to the social optimum, has certainly contributed to the restriction to the demand management discussions in some contexts like dense urban areas or environmentally sensitive areas. (cf. ECMT 2003b)

Congestion costs are "external" because individual decisions to use a road or any other crowded transport facility give rise to time delays for other users. This has led to the view that congestion is to be considered like other external costs. Behaviour leading to external costs should be contained by taxing their negative outcomes.

The conceptual literature following this view paid little attention to the informational function the price for infrastructure use has for decentralised infrastructure investment decisions, but focused strongly on demand management aspects (e.g. Proost *et al*., 2003b).

The early literature on optimal pricing and self-financing of infrastructure, taking account of congestion costs, made rather specific assumptions on the relationship between congestion costs and the increase of transport infrastructure capacity. Disregarding non-linear pricing or two-part tariffs, it assumed that there are "constant returns to scale in the congestion technology" (Small, 2005). That is, it was assumed that travel times remained unaffected when both the infrastructure use and the road capacity expanded continuously and in the same proportion. This, in turn implies that there are constant economies in road construction. The fact that most transport infrastructure can only be expanded in major discrete steps is given little importance in these considerations. (cf. the review in Verhoef, forthcoming)

Another root of the pricing discussion, of more relevance for the link between pricing and the capacity of infrastructure facilities, is the literature explaining to what extent the production of public goods can be decentralised (cf., e.g., Starrett, 1988, ch. 4). Public goods are facilities which can be utilised by more than one user, without one curtailing the usage of the other. Infrastructure goods belong to the particular class of these goods where "non-rivalry" in consumption is due to the physical indivisibility of the good. This implies that at low levels of demand, more users can enjoy the services without any reduction of others' consumption. Congestion is another term for the crowding of such public goods. What makes crowding or congestion interesting when thinking about public finance is the fact that it allows for a decentralised market-like supply of this type of public good. The collection of users is sometimes called a "club", to raise the connotation that supply should be organised as if the group of users would collectively decide how this should be done.

The substance of this argument is identical to the one in the preceding section on external costs of road damage. The "indivisibility" means that infrastructure facilities are associated with high fixed costs. The additional administrative and maintenance costs which can be attributed to an additional user are low and often considered to be constant. That is, without taking external costs into account, average costs are decreasing. Adding congestion costs to the picture, the social costs will at some usage level become higher than the core infrastructure costs and increase more than proportionately, as in Figure 3. When this is the case, there will be a minimum average cost level, which indicates how the unit of service should be priced and what size the facility should have. The level of congestion costs is positive at the point of minimal social costs. That is, infrastructure charging does not aim at the complete removal of congestion. Rather, beyond a use level of u*, the users favour the expansion of infrastructure capacity over reducing infrastructure use through increased prices. Taking both road damage and congestion into account will have the consequence of making the exponential growth of the social costs of infrastructure services stronger.

Other external costs

In recent pricing discussions, in particular in the European Union, other external costs, for example for environmental damage from transport and accident risks, have been included in the calculations for pricing prescriptions. This has to some extent to do with the focus on the demand management dimension of pricing; environmental damage due to pollution is seen as analogous to congestion costs. The environmental damage associated with using transport infrastructure should be priced to contain behaviour which incurs costs on others. Including all kinds of external costs in calculating prices for infrastructure use might have the advantage of saving costs for the fiscal administration. It has, however, important disadvantages for establishing a quasi-market for infrastructure service, linking a “fee for service” to supply decisions:

In general, external costs should be corrected for as close to their causes as possible. In the case of air pollution, pollutants should be taxed as a matter of principle. This would give the most direct and strongest incentives to avoid pollutants by changing transport behaviour or the technologies used. Such specific measures to correct for external costs might be expensive to implement and require less specific proxy instruments.

However, when infrastructure charges are designed to be used for infrastructure investment and/or maintenance, the attempt to internalise external costs by charging per unit of service leads to dysfunctional cost recovery mechanisms: if, for example, environmental costs are included in the calculation base, the funding of infrastructure is more generous the lower the environmental standards of the vehicle fleet

In the process of policy decision-making, road managers, the construction industry, etc., might thus have reason to oppose legislative measures to reduce the environmental damage caused by transport. The exclusion of environmental or other costs, from the infrastructure charging scheme requires other, more specific measures to reduce them. The need for specific corrective measures for specific external costs was discussed on the basis of the background paper by Santos (forthcoming), which also looks into the costs and benefits of earmarking corrective taxes. Measures to correct external costs, for example fuel taxes to reduce CO_2 emissions, will shift the demand curve in the above diagrams to the left. The above arguments on optimal pricing and capacity choice remain unaltered if “demand” is the demand that results after appropriate corrective measures (other than charges for infrastructure use) have been put in place.

INFRASTRUCTURE FUNDS AS COST RECOVERY MECHANISMS

To establish an argument that a quasi-market for infrastructure services is conceivable and feasible does not necessarily imply that it is desirable. One section of the Round Table discussion tried to clarify whether off-budget cost recovery mechanisms are needed, and what the advantages of such a mechanism would be compared to a traditional fiscal system of infrastructure investment and maintenance funding.

The Round Table discussion referred to the debate between the International Monetary Fund and the World Bank on the usefulness of road funds (mainly in developing countries). The IMF saw the establishment of infrastructure funds as a threat to an ordinary budget process, limiting the opportunities for rational fiscal programming, particularly in times of changing tax revenues. There was also the concern that infrastructure funds could and would be abused for other than infrastructure funding purposes or for ostentatious projects. In fact, the application of (some version of) optimal taxation concepts would result in the same outcome as the quasi-market sketched above (e.g. Diamond, 2003a): public goods should be provided by raising finance through fixed charges and linear taxes on net-trades in goods and services. The question is whether the outcome of the political process to provide transport infrastructure will be equivalent to a planning result that exclusively takes the infrastructure users' interests into account (Potter, forthcoming).

The World Bank's experience was in stark contrast to such an expectation (Gwilliam, forthcoming). Inadequate road maintenance is seen as a problem of the highest priority in many client countries. In Africa, it was estimated that during the two decades of the 70s and 80s, road stock to a value of US$45 billion was lost due to inadequate maintenance, which could have been avoided by expenditure on preventive maintenance of only US$12 billion (Brushett, 2002). A systematic comparison between planned and realised expenditures for transport investment and maintenance only exists in rare cases. A World Bank assessment of such information revealed that in all countries in the sample, actual expenditures were far below the planned expenditure levels. The highest value was 58 per cent and the lowest 15 per cent.

While the problems of funding transport infrastructure investment and maintenance may be less severe in countries with a highly developed fiscal administration, there is an almost universal effort to increase the number of public-private partnerships in transport infrastructure in order to mobilise funding for planned transport infrastructure projects. This suggests that there is the perception that the budget process favours other portfolios relative to transport policy. In these cases, off-budget cost-recovery mechanisms following the self-financing concept presented above would improve the contribution of the transport sector to the overall economic development.

In part as a response to the concern that infrastructure funds could be abused, to earn hidden profits or bureaucratic rents, the concept of "second generation road funds" has been developed. This concept gives an example of how to contain the risks of the diversion of funds and to ensure that they work to the advantage of infrastructure users. The infrastructure funds should be organised according to the following features:

i) Charges should be established in addition to and entirely independent of the determination of levels of taxes on road users for general revenue purposes.

ii) Charges should be directly transferred into the Fund, outside allocations from the general budget.

iii) The infrastructure fund should be managed by a board representing infrastructure users who would simultaneously determine the level of charge and the service.

iv) The users' board should decide efficient internal allocation procedures to determine day-to-day allocation decisions.

That is, the "second generation infrastructure funds" should have the status of an autonomous agency, controlling the funding of maintenance, and possibly investment. They should be directed mainly by the users to give a strong incentive to insist on commercially and professionally efficient management.

A critical discussion of infrastructure funds continues, however. This is mainly due to the fact that the discussion of institutional issues and good fund governance is detached from a discussion of where the resources to feed the fund should come from. In many countries, in particular those where the fiscal administration is weak, the parties supporting infrastructure funds propose to fund them by fuel taxes. As fuel taxes are levied for a variety of reasons some of which are largely unrelated to a national infrastructure policy this would violate the principle of the second generation funds. As fuel taxes reduce CO_2 emissions, feeding infrastructure funds by fuel taxes would lead to the undesirable association of the infrastructure quality with damage done to the environment. The often high level of fuel taxes would also raise doubts about the credibility of an infrastructure fund policy supporting concerns about bureaucratic or monopoly rent seeking.

The implementation of infrastructure funds should be strictly associated with the efficient rules to endow them set out above. This is also likely to increase the acceptability of a quasi-market for infrastructure services; all revenues from charging will be returned to supply. Capacity choices are made to provide the service at minimum average costs, including the external costs for road damage and congestion. Per-unit-of-service charges will be based on the extra costs from small increases in the level of usage. With such a model, users can more easily perceive the charge as a payment for present or future infrastructure services. It would avoid the impression existing in public discussions that infrastructure charges are tax increases by another name.

CONCLUSIONS

The current debate on infrastructure charging is dominated by demand management considerations, i.e. the question how transport can be contained or the modal split changed to better take account of, for example, environmental or accident costs associated with transport. In this discussion "prices" appear as a form of a tax. The Round Table had the objective to take this discussion a step further and find answers to the following questions:

- Is it possible to identify a charging scheme that would not only convey cost signals to the users of the transport system but also provide information for infrastructure managers on where to invest and which capacity to choose?

- Would the pricing rules discussed under the restriction to a demand management perspective remain in place when charged with the broader function of guiding infrastructure investment?

- Should a charging and cost recovery system entirely rely on a per unit of service (in many cases a per-km) charge?

- Should the charges for infrastructure services be considered a tax?

- How can a cost recovery mechanism be designed to work to the benefit of the user? How can risks of an abuse of a cost recovery mechanism like a infrastructure fund be minimized?

The Round Table arrived at the following answers:

Charges for infrastructure services can provide information on the location and the scale of infrastructure investment. To do so they have to follow strict pricing rules.

The charges per unit of infrastructure service should be based on the additional costs caused by the last additional unit of service for road administration, maintenance, and external costs which are directly associated with the provision of the services, i.e. vehicle operation costs that derive from a road damage externality and from congestion. The inclusion of other external costs at the margin, like environmental costs and accident costs, may lead to an under-utilisation of infrastructure facilities. Fiscal corrections of these external costs of transport have to use other instruments than per unit of service infrastructure charges.

Where the external costs of road damage and congestion costs are not high, i.e. in cases of relatively low levels of usage, revenues from pricing according to the additional costs for the last unit of service will not cover full costs. In such cases a fixed charge has to complement a per unit charge to install a mechanism to recover the full costs of infrastructure services.

A cost recovery mechanism which is separated from the budgetary process could help to correct for a disadvantageous position transport policy sometimes seems to have in the budgetary process. It would also avoid the impression to the taxpayer and the public at large that the introduction of a infrastructure charging scheme is just another tax increase.

The organisational design of a cost recovery mechanism like an infrastructure fund has to make sure that the resources made available to the fund are those identified by the above mentioned concepts of optimal pricing and investment. It should make sure that infrastructure users have control rights over the fund management.

REFERENCES

Brushett, S. (2005). Management and Finance of Roads in Africa. SSATP Discussion Paper No. 4, Road Management and Financing -RMF Series. World Bank Washington.

Commission of the European Communities (2001). White Paper: European Transport Policy for 2010. Time to Decide. Bruxelles.

Diamond, P. (2003). *Taxation, Incomplete Markets, and Social Security*. Cambridge, Mass.

ECMT (2003). *Reforming Transport Taxes*. Paris.

Gwilliam, K. (forthcoming). The role of road funds in improving maintenance. In OECD/ECMT Transport Research Centre (ed.), *Transport Infrastructure Charges and Capacity Choice*. Paris.

Kopp, A. (forthcoming). Fairness, efficiency and the simultaneity of pricing and infrastructure capacity choice. *European Transport* 26.

Newbery, D. M. G. (1988). Road damage externalities and road user charges. *Econometrica* 56: 295-316.

Potter, B. (forthcoming). Budgeting for road maintenance. In OECD/ECMT Transport Research Centre (ed.), *Transport Infrastructure Charges and Capacity Choice*. Paris.

Proost, S., K. Van Dender, C. Courcelle, B. L. De Borger, and J. Peirson (2002). How large is the gap between present and efficient prices in Europe? *Transport Policy* 9: 41-57.

Prudhomme, R., and J. P. Bocarejo (2005). The London congestion charge: a tentative economic appraisal. *Transport Policy* 12: 279-87.

Santos, G. (forthcoming). Road user charges and infrastructure. In OECD/ECMT Transport Research Centre (ed.), *Transport Infrastructure Charges and Capacity Choice. Round Table Report 136*. Paris.

Small, K., and Y. Jan (2001). The value of "value pricing": second-best pricing and product differentiation. *Journal of Urban Economics* 49: 310-336.

Starrett, D. A. (1988). *Foundations of Public Economics*. Cambridge, Mass.

Verhoef, E. (forthcoming). Self-financing of infrastructure services. In OECD/ECMT Transport Research Centre (ed.), *Transport Infrastructure Charges and Capacity Choice. Round Table Report 136*. Paris.

APPENDIX 2

CONCLUSIONS OF ROUND TABLE 136: ESTIMATION AND EVALUATION OF TRANSPORT COSTS

INTRODUCTION

Round Table 136 on the “Estimation and Evaluation of Transport Costs” was held in Paris on 1-2 December 2005. The Round Table discussion was based on four background papers, introducing the substantive issues of the topic.

Antonio Estache (World Bank+) and Lourdes Trujillo (Universidad de Las Palmas de Gran Canaria) contributed a paper discussing the importance of transport cost estimation to evaluate transport policies and reviewing the most important methods. Piet Rietveld *et al*. presented a paper on the estimation of costs of the different transport modes for setting regulatory limits to infrastructure charges. Carlos Barros’ paper set out an econometric approach to measuring the efficiency of airports and applied it to Portuguese airports. Finally, Philippe Gagnepain (Universidad Carlos III, Madrid) and Marc Ivaldi (Université de Toulouse) extended this econometric approach to the estimation of cost functions by taking into account the effect of incentives of regulatory regimes on the efforts of public transit firms and airlines to reduce their costs.

The Round Table was chaired by Tae Oum (University of British Columbia), a leading researcher on costs studies, and on transport policies for the aviation sector in particular. Other leading experts from Europe, Japan and the United States of America, with backgrounds in engineering as well as economic approaches to the topic, participated in the Round Table. The background papers and the list of participants have been distributed to the Committee members.

The motivation for the Round Table on the estimation and evaluation of transport costs is related to a number of current transport policy objectives:

-- Despite the almost universal commitment of transport policy to achieve efficiency in the transport sector, the empirical basis to assess the productivity of the transport sector overall, either in the provision of infrastructure services or the operation of individual transport modes is often weak. Relevant data are not available or they lack credibility due to their poor quality.

-- A major impetus for carrying out cost studies currently results from efforts to prescribe pricing rules to public or private providers of transport services. To do so, information is required on resource requirements for the technically efficient provision of infrastructure services, i.e. the least cost per unit of additional service of a certain quality. These costs might differ from currently observable costs as the lack of competition might allow for administrative rents or ostentatious projects.

-- More generally, due to the fact that not all transport operations and infrastructure facilities can be exposed to competition in the market (Knieps, 2005), regulators critically depend on information on costs and cost functions to determine the scope and substance of regulation. Yardstick competition (Bouf and Leveque, 2005) rewards transport and infrastructure firms for being close to the best practice technologies and therefore requires the empirical identification of the best practice cases.

-- As even firms that have the lowest level of costs in practice might fail to adopt available blueprints or inventions, observed costs might substantially differ from an implementable lower level of costs. In these cases up-to-date engineering information is essential for taking decisions on pricing and regulation, including the provision of incentives to adopt new technologies. As local prices of inputs or different skill classes of labour will in many cases differ, the combination of economic or econometric analysis will have to complement engineering information in determining the least costs of transport or infrastructure services. The Round Table therefore also had the objective of contributing to the communication between experts having different professional perspectives on the assessment of costs and efficiency in the transport sector.

Following the discussion plan of the Chairman, and based on the background papers, the Round Table discussed the following sub-topics:

- Following the presentation of Antonio Estache and Lourdes Trujillo (forthcoming) the pros and cons of the different methods to estimate and evaluate transport costs were looked into. The usefulness of the different methods and their link to transport policy objectives depend on the purpose of the cost analysis and the costs of their application.

- As a concrete example of the difficulties that follow from policy objectives for cost studies including the estimation of social costs, and the problems that arise in practice from the lack of data and resources available, the Round Table discussed a proposal for a charging mechanism to recover maintenance costs in all land transport modes in the Netherlands (Rietveld *et al*., forthcoming).

- As an application of a formal cost benchmarking analysis, the participants discussed the comparative analysis of the efficiency of Portuguese airports, presented by Barros (forthcoming).

- This type of analysis was extended to include policy variables which reflect the regulatory regime in force, as in the paper of Gagnepain and Ivaldi (forthcoming). Its merits were discussed on the basis of the application to public transport regimes in French cities and the liberalisation of European airlines.

COST INFORMATION AND TRANSPORT POLICY OBJECTIVES

Perhaps the most general objective of transport policymakers is to achieve an efficient transport sector, or a transport sector that contributes as much as possible to overall economic development. This implies in other words that transport policy is aimed at or should be aiming at achieving mobility for passengers and goods with a minimum sacrifice of other goods and services, or a minimum value of inputs.

More specifically, higher than minimum transport costs reduce the competitiveness of national economies by increasing the prices of imported goods, reducing the net returns from exporting goods and reducing real incomes by increasing the prices of domestic products (e.g. Clarke, Dollar and Micco, 2004).

On the subnational level, high transport costs distort the division of labour between different regions and remove part of the agglomeration benefits of cities and metropolitan areas. These agglomeration effects are considered to be central to knowledge accumulation and long-run economic growth (Black and Henderson, 1999).

On the local level, high transport costs impede the functioning of goods and factor markets. An example of such a malfunctioning of markets that has recently received much attention in research is the spatial mismatch in labour markets, resulting from the high costs of intra-urban passenger transport (Patacchini and Zenou, 2005).

In general, a reduction in production costs increases the real income of the population. Pressures to reduce costs are normally brought about by market forces. However, major parts of the *transport* sector are not subject to the market forces that guide firms and households to adapt to the supply of inputs and to adopt least-cost technologies (Round Table Report 129). In this context, transport policy measures are required to signal relative scarcities to producers and consumers and to help the implementation of least-cost solutions in providing infrastructure and transport services. These incentives can only be provided if the costs of infrastructure services and transport operations are assessed and monitored.

To effectively carry out regulatory or pricing policies, transport policymakers depend decisively on reliable and timely cost information. The fact that the data will be used to more effectively regulate firms which are at the same time data providers may lead to resistance against data collection. In these cases data collection may be associated with high costs to monitor its quality. Moreover, data on actual costs of infrastructure facilities and transport services may be rather uninformative with respect to the overall efficiency objectives. If providers enjoy monopolistic powers through an essential facility, some of the costs may reflect monopoly rents. These may result from overstaffing or higher salaries than those paid for similar jobs in the private industry. In other cases service providers may have few difficulties in receiving government subsidies to cover deficits. The reported cost levels might then, for example, reflect a high degree of unionisation of the workforce which can be used to negotiate high wages and high levels of employment. (cf. Laffont and Tirole, 1986). As a first step, measuring cost levels which are observable in the market allows the identification of best practice infrastructure facilities or transport firms. Knowing about best practice firms helps to reduce relative inefficiencies in the market by providing information and/or regulatory incentives to imitate best practice technologies and behaviours.

Even accepting this reduced ambition, it is not obvious how costs should be measured and what are the most useful accounting methods. Regulatory purposes are helped substantially by establishing separate cost accounting entities for those parts of transport activities or infrastructure services which are or can be exposed to market pressure to reduce costs (Knieps, 2005) and those which enjoy the powers of a natural monopoly or an essential facility (Estache and Trujillo, forthcoming). Such a separation would reduce the opportunities to diffuse regulatory pressures by arbitrarily assigning cost items to activities.

Discussions on the relative performance of transport firms or infrastructure facilities suffer from the fact that there is no unique way to measure performance. If partial performance measures are used there are naturally a number of indicators which might give conflicting signals on performance.

Table 1. **Partial Performance Indicators**

Operator	Labour (L)	Capital (K)	Output (Y)	Y/L	Y/K
A	200	2	2000	10	1000
B	400	1	2000	5	2000

Estache/Trujillo, forthcoming.

Looking at labour productivity in column 5, operator A would rank higher. Choosing capital productivity as the ranking index, operator B would perform better than operator A. The ambiguity of the partial performance measure can only be resolved if output per "bundle" of inputs is measured. What is to be considered the reference bundle for performance measurement can, unfortunately, be defined in different ways, too. The measurement of *total* performance of a firm or facility requires the use of index numbers to express total factor productivity (Coelli *et al*. 1998, ch. 4). In most empirical applications, the Toernquist Index formula is used for purposes of output and input index calculations (see Annex). For this index, as for most other applications when aggregating the influence of individual inputs on the output, a functional relationship is assumed which implies that a certain percentage increase in individual inputs leads to a constant percentage increase in output.

Measuring the efficiency of a facility or firm has three dimensions which the formal cost estimation methods try to disentangle (see Annex).

- Is the least-cost technology used? That is, are there maybe other technologies that require smaller amounts of one or more inputs per unit of output? Answers to these questions determine whether the firm is *technically efficient.*

- Does the provider adequately respond to the prices of the inputs needed? In other words, is the right combination of inputs used? *Allocative efficiency* is achieved if no cost savings could be made by changing the shares of the different inputs in total costs.

- *Scale efficiency* is achieved if the costs per unit of service cannot be reduced by changing the size of the firm or the facility.

The Round Table discussed to what extent the standard production function and cost function estimation methods neglected engineering information. A research program on engineering production functions in economics (cf. the review in Wibe, 1984, and Chenery, 1992) is currently inactive. It had very little impact on cost estimation and evaluation despite remarkable applications to the transport sector. Vernon Smith (1957) applied the engineering production function approach to the trucking industry, De Salvo (1969) to the railway industry and Hildenbrand (1981) to the Norwegian tanker fleet. That engineering data can be usefully integrated into economic simulation models to help regulators has recently been demonstrated for the telecommunications sector (Gasmi *et al.,* 2002).

Cost estimation in practice

Even if it is clear in principle which cost information is needed and how cost estimations should be done, applications to solve concrete policy problems often look different. The reasons for such differences are many.

- First of all, shortage of cost data seems to be a problem that holds in particular for the transport sector.

- As mentioned before, the data quality might be imperfect in cases where primary data providers have their own transport policy or economic interests connected to the reported data.

- The data collection costs or the costs for monitoring the data may be such that the benefits for the policy decision-making process do not justify the expenses.

- Finally, the application of cost estimation and cost evaluation methods may appear difficult to communicate in the policy process and imperfect surrogate methods are applied instead.

However, the imprecision of quantitative information obtained from employing rough, ad hoc measures may be high. The Round Table discussed these kinds of problem, based on the background paper of Rietveld *et al.* on the introduction of infrastructure charges for road, rail and inland navigation to recover infrastructure maintenance costs in the Netherlands.

The starting point for the proposal is the concept of social marginal cost pricing. To charge users of infrastructure according to the additional costs which are caused by the increase in infrastructure usage is the pricing rule that would induce the optimal use of infrastructure facilities. A price higher than marginal costs would induce an underutilisation of the facility and lower prices an overuse, with high levels of congestion. It is acknowledged that in principle not only additional infrastructure costs should be charged, but also external costs in the form of environmental damage, increased safety risks, increased congestion costs, etc. at the margin. These concepts for pricing of infrastructure services are, however, unrelated to cost recovery. In general, a self-financing charging policy would require a marginal cost pricing per unit of service plus a fixed access charge to cover the full costs. Moreover, incentives to contain external costs of transport are not necessarily best provided by charges per unit of infrastructure use. Other instruments to internalise them are in place (like fuel taxes to contain CO_2 emissions) which would have to be revised should charging for external costs be transferred to charges for infrastructure use to cover maintenance costs.

Against the backdrop of these difficulties, a relatively simple model of cost allocation has been proposed. It starts from the observation that not all maintenance costs vary with infrastructure use and that there are hence fixed and variable maintenance costs. For the cost recovery scheme, both the variable and fixed infrastructure costs are allocated to different vehicle types, leading to recommendations on charges per distance unit.

Social transport costs and cost functions

The core marginal costs are identified by engineering information. For roads, they are computed on the basis of axle load factors, which are calculated using the AASHO 4^{th} Power Rule. Similar engineering information has been used to calculate marginal infrastructure costs for rail and inland waterways. Per vehicle-km charges to cover extra per vehicle-km costs in terms of wear and tear, noise and increased accident risks will not, however, cover the full average costs per vehicle-km. This holds in particular, if congestion costs -- the external costs directly linked to infrastructure use -- are disregarded in the price calculation. Excluding other, non-linear pricing schemes, fixed costs have to be allocated to vehicle types and to be translated into per-km charges.

As is shown in the cost figures of the background papers, even if we restrict the cost recovery scheme to maintenance costs, a substantial share of the costs are fixed costs, for example 55 per cent of the Dutch highway maintenance costs are fixed costs.

The allocation of fixed costs to per vehicle-km prices often follows ad hoc rules, referring to technical characteristics of vehicles and/or external costs which are expected to be associated with certain vehicle types. The Dutch proposal for the recovery of maintenance costs allocates fixed maintenance costs to vehicle types according to average vehicle sizes. By doing so, the proposal follows the recommendation of the European Commission to impute fixed costs on the basis of kilometre-equivalence factors (EU 1999, Directive 1999/62/EG of the European Parliament and the Council concerned with the Charging of the Use of Certain Infrastructure Services to Heavy Goods Vehicles, European Communities, Brussels). The base rate calculated in this way is supplemented by differentiating the charges to cover the fixed maintenance costs according to the social external costs of transport, in particular the contribution of vehicles of different classes to environmental damage.

The Round Table discussion revealed the problems of partial cost recovery schemes, with the restriction of cost recovery to maintenance costs as a particular example. If two different modes had identical total costs but different shares of construction and maintenance costs, basing pricing decisions on maintenance costs only discriminates against the maintenance-intensive mode. The cost estimates presented for a maintenance cost recovery scheme suggest that charges per rail passenger-km or rail ton-km would be far higher than for roads. This result might or might not be confirmed by an analysis of total costs (including construction costs). Translating fixed costs into per unit charges for services would in general conflict with a rational use of infrastructure. It could be inferior to a (two-part) combination of fixed access fees and marginal cost-based per-km charges. (cf. Round Table 135). For non-congested infrastructure facilities, full cost pricing will in general lead to an overpricing and underutilisation of the infrastructure.

Moreover, over- or underutilisation of infrastructure can easily be induced by accounting for external costs of transport in an ad hoc manner. Of particular importance for the modal split are congestion costs which should be an integral part of pricing schemes. Imprecise estimation methods -- like basing the estimate of additional congestion costs just on the size of a vehicle which is then used for fixed maintenance cost distribution to vehicle types -- are likely to lead to an underpricing of the use of congested facilities. They might contribute to a failure to achieve congestion reduction targets and, by overpricing of uncongested facilities, to an underutilisation of uncongested infrastructure.

A more complete collection of data, a precise estimation of costs and cost functions and the implementation of more complete cost allocation schemes may pay off by helping the implementation of infrastructure capacities and inducing use levels that lead to prices closer to the minimum costs per unit of infrastructure service.

Barros' paper proposed to measure the efficiency of Portuguese airports by means of a stochastic cost frontier, presented in detail in the Annex. The stochastic cost frontier estimation aims at the identification of the minimum costs for all output levels, taking into consideration that there might be random measurement errors or omitted variables. The estimation model extends the standard approach set out in the Annex by accounting for "inefficiency effects", effects that influence the distribution of the error term of the estimation function.

The most important result of the exercise is a strong indication of the size advantages of airports. Given the public concern about the fiscal burden arising from the deficits of airports, the result might suggest an indication of airport overcapacity. For conclusions on national airport policies, further analysis of airport operations might be required. The standard cost frontier estimation does not take into account that different airports might serve different sub-markets in a differentiated overall market for airline services. For example, network effects make it desirable for international carriers to have a single, large national hub. The scale effects shown in the cost estimation may, however, not be replicable for other airports. Moreover, the expected increase in the specialisation of airports (cargo, low-cost carriers) might lead to decreased average costs, or the increased productivity of smaller airports in the future.

It might appear tempting to conclude from the results that a reduction in the number of airports and an increase in their average size should in general lead to efficiency gains. This would disregard congestion costs which form an important cost component of modal infrastructure facilities. Adding congestion costs to the core infrastructure costs would fundamentally change the cost function of infrastructure facilities like airports. Instead of a cost function which implies continuously falling costs per unit of service, a u-shaped cost curve would result, with the increasing congestion costs of additional traffic leading to increasing total average costs. The minimum costs per unit of airport service could serve as an indicator of the desirable airport capacity. The results of the study on the costs of Portuguese Airports confirm the need to include external costs of infrastructure facilities, in particular congestion costs, to identify optimal airport sizes.

Cost function estimation and regulatory regimes

Gagnepain and Ivaldi (forthcoming) extend the standard stochastic cost function approach by introducing information on regulatory regimes and the associated incentives to reduce costs in the estimation approach. The differences in the incentives of the regulatory regimes are due to differences in dealing with asymmetric information between firms and regulators. Their approach is applied to a study of public transport in France and to the deregulation of European airlines.

Thirteen urban areas had been included in the study of 59 networks, with data collected over the period 1985 to 1993. In these areas, urban public authorities are in charge of regulating the transit system, whose services are provided by a single operator. Two regulatory regimes are observed in practice. First, under a cost-plus contract, all costs of the public transit firm are reimbursed *ex post*. As such a scheme provides at best very weak incentives to reduce costs, the estimation approach assumed that in such a context firms do not try to cut costs. The second regulatory regime is a fixed price contract, leaving the operator with the responsibility for insufficient revenues and cost overruns. As 60 per cent of the costs of public transit are labour costs, the cost reduction effort will mainly consist of training drivers, organising work to avoid conflict, etc.

Results on two-thirds of the sample clearly confirm expectations: A first group with the highest productivity operates under a fixed-price contract, suggesting that the regulatory regime has a decisive influence on the efforts to reduce costs. The vast majority of a group of 20 operators in the middle range of productivity operate under a cost-plus regulatory regime, in line with the expectation that firms whose costs are reimbursed take less action to reduce costs. In contrast to this supposedly clear picture, the productivity of a third group, with the worst performance, contains firms operating under both regulatory regimes. It shows that other determinants than the regulatory regime can play a role for the transport providers' performance. The result can, for example, be interpreted as indicating that some cities suffer from infrastructural and institutional legacies which exclude the fact that fixed-price regulation has the same productivity-enhancing effects it had for the best performers.

On the liberalisation of the European airline industry, alternative scenarios are compared concerning deregulatory packages introduced by the European Commission:

a) firms realise efficiency independently of de-regulation; i.e. the estimation model does not need an effort and inefficiency term;
b) firms are inefficient but do not react to deregulation by cutting costs;
c) firms start efforts to cut costs after the implementation of the third EU package in 1992;
d) deregulation changes the behaviour of firms affected by the introduction of bilateral agreements (British Airways, KLM, Lufthansa and Sabena after 1985, and other airlines in 1993).

The cost estimation, which translates the different scenarios into different specifications of the net inefficiency term, allows the testing of these scenarios against each other with clear implications for regulatory policies.

-- Scenario (d) is rejected in favour of scenario (c). This suggests that the third aviation package was far more effective than bilateral agreements in affecting the cost reduction efforts of the airlines.
-- Scenario (a) rejected in favour of scenario (c): that is, the inclusion of an inefficiency term and, accounting for the possibility that deregulation changed the airlines' cost management, led to superior estimation results. As (a) is a standard model for studying firms' behaviour in oligopolistic markets, the outcome of the test suggests that the outcomes of the standard approach have to be read with caution.

-- The standard approach of cost estimation suggests that the European airline industry is characterised by increasing returns to scale. By contrast, Scenario (c), which fared better in the tests, led to the result that it is characterised by constant returns to scale.

Overall, the results confirm the positive impact of airline deregulation on the adoption of cost-reducing technologies and its effectiveness in inducing efforts to increase labour productivity.

To obtain a better understanding of the potential for cost reductions, and the implications this has for the market structure, combining engineering information and econometric cost estimations to construct simulation models for transport subsectors may be helpful. Research along these lines has so far been restricted to the market for telecommunications services (Gasmi *et al*., 1999, 2002).

Conclusions

Without information about cost levels in transport or infrastructure service provision it is impossible to identify resource requirements for transport policy. Knowledge of best practice technologies and their associated costs provides the basis for setting operational productivity targets. As was discussed in Round Table 129, data collection and cost estimation efforts are, for example, essential for the implementation of new regulatory mechanisms such as yardstick competition. To help these fundamental planning functions, the Round Table addressed the following questions:

- Do planners and researchers need a broader statistical data base to help transport policymaking?
- Are the methods used in practice adequate to generate the information required for policymaking?
- What are the limits of currently available cost assessment methods? What should guide the selection of methods?
- Are there deficiencies in currently available approaches to estimating and evaluating costs, with a need to develop them further?

Based on the background papers, the Round Table discussion arrived at the following answers:

- There was a consensus among Round Table participants that there are deficiencies in the statistical information available, identified for example by comparison with the data available for other infrastructure sectors. Current resource allocations to transport data collection in some member countries leads to the concern that the situation is not improving.

- Decisions on the scope and design of data collection efforts face a number of problems. Extending the scope of data collection requires a careful estimation of the benefits for transport planning, and an answer to the question of whether they justify the additional costs. Data quality problems arise from the informational asymmetries between data users and data providers. At least in some instances the anticipation of data use, for example for regulatory purposes, will invite the provision of distorted data. When this concerns technical information, engineering data may help to check the data quality.

- Cost studies often take inadequate account of external costs. To avoid an ad hoc inclusion or exclusion of external cost data, a careful theoretical analysis of the policy problem at hand must precede the cost estimation. This would also avoid the particularly high costs of an analysis of types of external cost, which might be irrelevant for the policy decision. For example, road damage externality and congestion costs are external costs which are essential for road pricing decisions while other types of external costs are not.

- The lack of data might make it difficult to avoid the application of ad hoc methods to assess costs in practice. As examples on infrastructure cost recovery mechanisms showed, the application of crude cost estimation approaches is likely to have negative effects on the design of policy measures, leading to wrong incentives and unintended policy consequences.

- The Round Table discussed the pros and cons of partial cost indicators, index numbers of overall factor productivity and methods to estimate cost functions. The estimation of cost functions allows the disentangling of the technical, factor price and scale dimensions of costs. However, as going from partial to general cost statements implies increasing costs for the analysis, a critical assessment of which information is necessary for which policy or planning problem is required. Participants reported that policy discussions often suffered from reading more into simple cost indicators than justified.

- Clear progress has been made in cost assessment methods by including asymmetric information and the consequent incentive structures in the cost estimation approaches. Their extension by including engineering information could help to address the fact that least-cost technologies may not be observable in the market and address the above-mentioned data quality problems due to the self-interest of data providers. An example of the feasibility of such an extension exists for the telecommunications sector.

ANNEX

Sketch of Cost Estimation Instruments

Index numbers try to avoid the deficiencies of partial cost indices. They aim at indicating the ratio between output variables and a bundle of inputs. The most frequently used productivity index is the Toernquist index.

The Toernquist total factor productivity index is defined in its simpler logarithmic form, comparing to entities s and t

$$\ln TFP_{st} = \ln \frac{OutputIndex_{st}}{InputIndex_{st}} = \ln Output\,Index_{st} - \ln Input\,Index_{st}$$
$$= \frac{1}{2}\sum_{i=1}^{N}(\omega_{it} + \omega_{is})(\ln y_{it} - \ln y_{is}) - \frac{1}{2}\sum_{j=1}^{K}(\upsilon_{jt} + \upsilon_{js})(\ln x_{jt} - \ln x_{js}),$$

where y denotes outputs, indexed by i, and x denotes inputs, indexed by j[4/]. The ω and υ denote the shares of goods i in total real output, and the shares of inputs j in total costs.

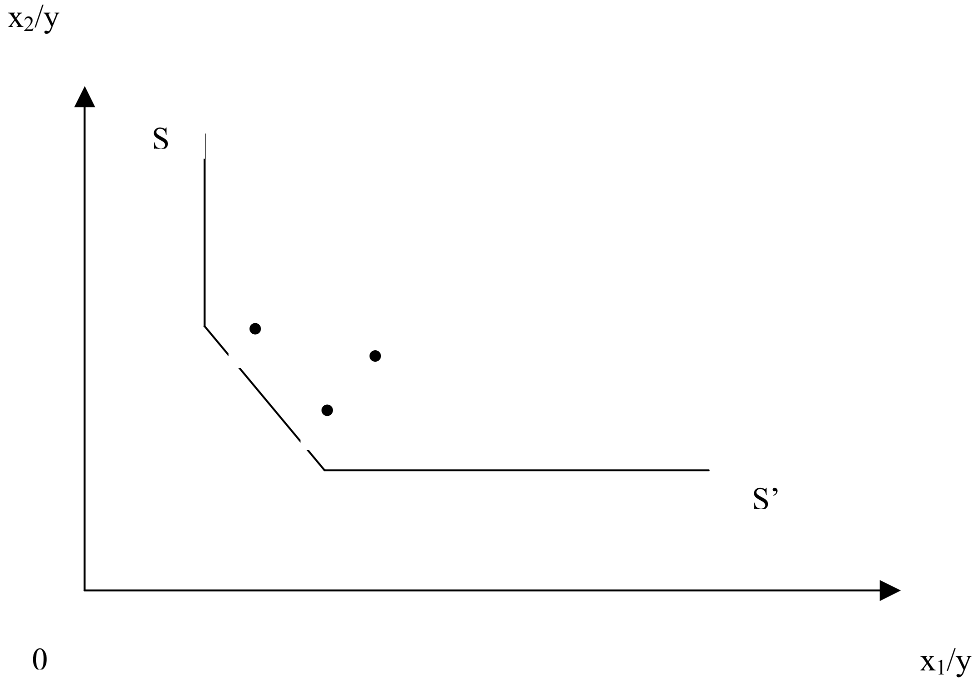
Total factor productivity indices indicate the overall cost effectiveness of projects or firms without, however, giving any indication about its sources. Of particular interest is whether inefficiencies are due to the use of technologies other than least-cost (technical efficiency) and/or whether providers fail to respond to the appropriate price signals (allocative efficiency). Moreover, inefficiency can result from mistakes in the choice of the capacity of firms or infrastructure facilities. In general, infrastructure capacity, and oftentimes transport operations, can only be changed in large, discrete steps. An important dimension of average cost levels in transport is therefore scale economies.

Indivisibilities and network economies which beset capacity choice in transport imply that average costs are decreasing for some potential levels of operation. To achieve scale efficiency, i.e. to choose the size of a firm or infrastructure facility such that it operates close to the minimum average costs, requires a precise forecasting of demand and the possible degree of congestion.

To make empirical statements about the different aspects of efficiency, the quantitative relationship between inputs and output(s) of the *fully efficient* firm or facility must be known. Such a production function is not known in practice. The bulk of the literature on the identification of an empirical production function follows the suggestion in Farrel's (1957) seminal article, to estimate such a function from sample data using either a non-parametric piece-wise linear technology or a parametric function. The first suggestion has developed into the Data Envelopment Analysis. The Data Envelopment Analysis tries to identify the set of minimal combinations of inputs required to produce one unit of service (the unit isoquant) by linear programming techniques, as illustrated by Figure 1 (Charnes *et al*., 1995).

4/ There are other index numbers which can be used for productivity measurement, which differ by certain mathematical properties. Which index number should be used depends on the purpose of the productivity study and the characteristics of the index. The discussion on ideal index numbers has been inconclusive (Diewert, 1992).

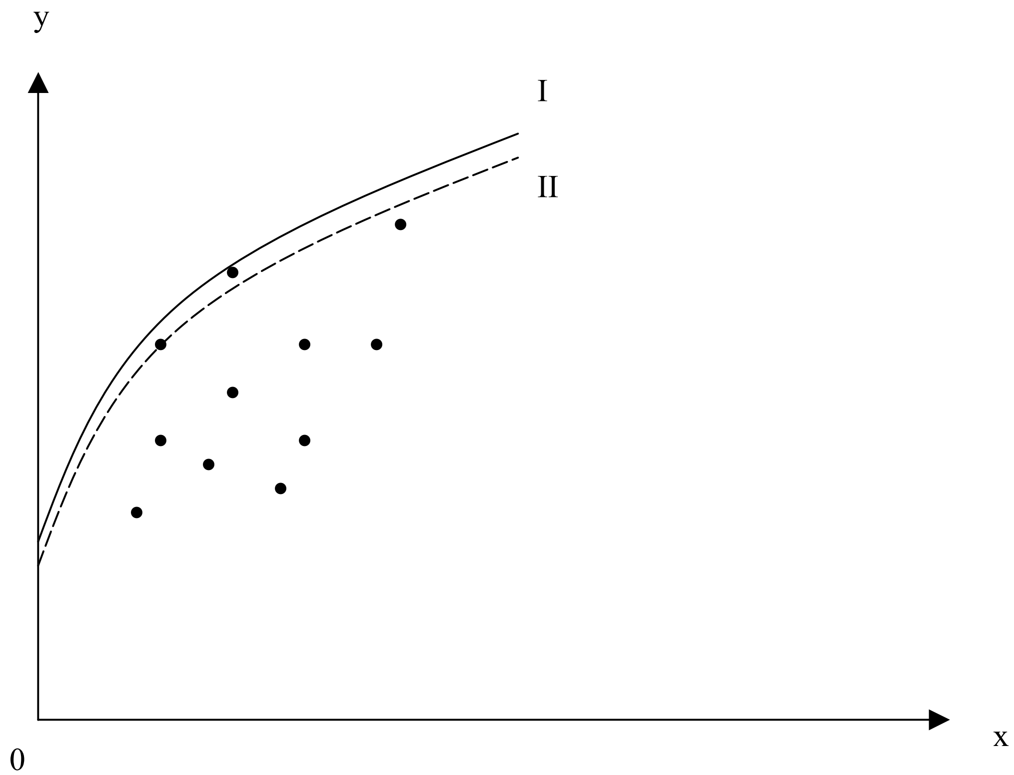
Figure 1. Piece-wise linear convex unit isoquant in case of two inputs and one output



All combinations of the inputs x_1/y and x_2/y which lie northeast of the convex hull SS' are inefficient. One unit of the service could be supplied with a smaller (at least one) input.

The approach used to identify the efficient relationship between inputs and output, implying the duality of minimum costs, is the frontier estimation approach (Färe *et al*., 1994). The frontier estimation approach tries to distil from observed input and output figures the least-cost values in the market. To illustrate the discussion on what is actually estimated, Figure 2 assumes that there is only one input x and one output y.

Figure 2. Estimation of a frontier cost function



Originally, the approach to determine a parametric production function started out from the following model:

$$\ln(y_i) = \ln(x_i)\beta - u_i, \text{ with i = 1,..., N and } u_i \geq 0, \text{ (graph I)}$$

where i indexes firms x_i is a (K+1)-row vector of the inputs of firm i, y_i the (scalar) output of the firm, β the (K+1) of parameters to be estimated and u_i a non-negative variable, indicating the inefficiency at the firm level. Aigner and Chu (1968) proposed in their pioneering article to determine the parameters by linear or quadratic programming, minimising the sum of absolute residuals u_i or the sum of squared residuals, respectively. This deterministic frontier model was criticised for being unable to take account of measurement errors, omitted variables or unpredictable, random human responses (Schmidt, 1976).

As a response to this criticism, Aigner, Lovell and Schmidt (1977) as well as Meeusen and van den Broeck (1977) independently proposed the stochastic frontier production function, adding a random error ε_i

$$\ln(y_i) = \ln(x_i)\beta - u_i + \varepsilon_i, \text{ with i = 1,..., N and } u_i \geq 0, \text{ (graph II)}$$

The ε_i's are assumed to be independent and identically distributed normal random variables with a mean zero and a constant variance. The inefficiency variables, required to be non-negative, are assumed to be exponential or half-normal variables. The stochastic frontier model permits the estimation of standard errors and hypothesis testing, using maximum likelihood methods, which was not possible with the early deterministic models.

In the background paper of Gagnepain and Ivaldi (2005) the level of inefficiency depends on the effort of the management of firm i, denoted by e_i, to produce efficiently, i.e. the effort to reduce the u_i's.

With the effort level included in the estimation equation, we have:

$$\ln(y_i) = \ln(x_i)\beta + \varepsilon_i + g(e_i - u_i), \text{ with i = 1,..., N and } u_i \geq 0.$$

The estimation of a cost frontier is obtained from minimising expenses for variable inputs for a given output level. The cost frontier is then a function of input prices, the output level, the inefficiency level and the effort to reduce the inefficiency.

BIBLIOGRAPHY

Aigner, D.J. and S.F. Chu (1968). On estimating the industry production function. *American Economic Review* 58: 826-839.

Aigner, D.J., C.A.K. Lovell and P. Schmidt (1977). Formulation and estimation of stochastic frontier production models. *Journal of Econometrics* 6: 21-37.

Barros, C.P. (forthcoming). Technical efficiency in the Portuguese airports with a stochastic cost function. In OECD/ECMT Transport Research Centre (ed.), *Estimation and Evaluation of Transport Costs. Round Table 136*. Paris.

Black, D. and V. Henderson (1999). A theory of urban growth. *Journal of Political Economy* 107: 252-84.

Bouf, D. and J. Leveque (2005). Yardstick competition for transport infrastructure services. In OECD/ECMT Transport Research Centre (ed.), *The Limits of (De-) Regulation of Transport Infrastructure Services. Round Table 129*. Paris.

Charnes, A., W.W. Cooper, A.Y. Lewin and L.M. Seiford (1995). *Data Envelopment Analysis: Theory, Methodology and Applications*. Boston.

Chenery, H.B. (1992). From engineering to economics. *Banca Nazionale del Lavoro Quarterly Review* (183): 369-405.

Clark, X., D. Dollar and A. Micco (2004). Port efficiency, maritime transport costs and bilateral trade. *Journal of Development Economics* 75: 417-450.

Coelli, T., D.S.P. Rao and G.E. Battese (1998). *An Introduction to Efficiency and Productivity Analysis*. Dordrecht.

De Salvo, J.S. (1969). A process function for rail linehaul operations. *Journal of Transport, Economics and Policy* (3-27).

Diewert, W.E. (1992). Fisher ideal output, input and productivity indexes revisited. *Journal of Productivity Analysis* 3: 211-48.

Estache, A. and L. Trujillo (forthcoming). Transport cost levels, productivity and efficiency measures: some theory and main policy conclusions. In OECD/ECMT Transport Research Centre (ed.), *Estimation and Evaluation of Transport Costs. Round Table 136*. Paris.

Farrell, M.J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society. Series A* 120, Part 3: 253-290.

Färe, R., S. Grosskopf and C.A.K. Lovell (1994). *Production Frontiers*. Cambridge, Mass.

Gagnepain, P. and M. Ivaldi (forthcoming). Measuring inefficiencies in transport systems: between technology and incentives. In OECD/ECMT Transport Research Centre (ed.), *Estimation and Evaluation of Transport Costs. Round Table 136*. Paris.

Gasmi, F., D.M. Kennet, J.J. Laffont and W.W. Sharkey (2002). *Cost proxy models and telecommunicaitons policy*. Cambridge, Mass.

Gasmi, F., J.J. Laffont and W.W. Sharkey (1999). The natural monopoly test reconsidered: An engineering process-based approach to empirical analysis in telecommunications. *International Journal of Industrial Organisation* 20: 43559.

Hildenbrand, W. (1981). Short run production functions based on microdata. In R.J. Aumann, J.C. Harsanyi, W. Hildenbrand, M. Maschler, M.A. Perles, J. Rosenmüller, R. Selten, M. Shubik and G. L. Thompson (ed.), *Essays in Game Theory and Mathematical Economics in Honor of Oskar Morgenstern*. Mannheim.

Knieps, G. (2005). The limits of competition in transport markets. In OECD/ECMT Transport Research Centre (ed.), *The Limits of (De-) Regulation of Transport Infrastructure Services. Round Table 129*. Paris.

Laffont, J.-J. and J. Tirole (1986). Using cost observation to regulate firms. *Journal of Political Economy* 94: 614-41.

Meeusen, W. and J. Van den Broeck (1977). Efficiency estimation from Cobb-Douglas production functions with composed error. *International Economic Review* 18: 435-44.

Patacchini, E. and Y. Zenou (2005). Spatial mismatch, transport mode and search decisions in England. *Journal of Urban Economics* 58: 62-90.

Rietveld, P., F.R. Bruinsma and M.K. Koetse (forthcoming). Infrastructure maintenance costs: a comparison of road, rail and inland navigation and implications for user charges. In OECD/ECMT Transport Research Centre (ed.), *Estimation and Evaluation of Transport Costs. Round Table 136*. Paris.

Schmidt, P. (1976). On the statistical estimation of parametric frontier production functions. *Review of Economics and Statistics*: 238-239.

Smith, V.L. (1957). Engineering data and statistical techniques in the analysis of production and technological change: Fuel requirements in the trucking industry. *Econometrica* 25: 281-301.

Wibe, S. (1984). Engineering production functions: a survey. *Economica* 51: 401-11.

APPENDIX 3

CONCLUSIONS OF ROUND TABLE 137: TRANSPORT, URBAN FORM AND ECONOMIC GROWTH

INTRODUCTION

Round Table 137 was the first Round Table to be held in a non-European country. It was hosted by the Institute of Transportation Studies of the University of California at Berkeley. The Round Table was chaired by Marty Wachs (RAND, Los Angeles). Background papers were provided by David Banister (University of Oxford), Elisabeth Deakin, (UC Berkeley), Gilles Duranton (University of Toronto) and Matthew Kahn (Tufts University).

Transport technology and the associated transport costs have always been among the dominant determinants of urban location and form. In the first half of the 19th century, most cities were tied to waterways, developing around harbours and by rivers and canals. Towards the end of that century, railways competed with inland waterways, and urban growth and form became determined by investments in rail terminals and by their scale economies providing advantages of proximity.

The high cost of intra-urban transport by horse and wagon favoured the creation of single manufacturing districts, located near harbours or railheads, with residential areas surrounding them. Before the advent of horse-drawn and electric street-cars, personal transport was mainly carried out by foot and horse-drawn carriage, implying a strong need to live close to the city centre.

With street-car transport, residential areas spread out around stations or street-car lines. The urban structure changed to a compact production core surrounded by residential areas, which were determined by mass transport facilities.

Only by the middle of the 20th century did the private car start to compete successfully with mass transit -- despite transit fares remaining flat in nominal terms (Barrett, 1983) -- by providing speed, privacy and convenience and being facilitated by the expansion and upgrading of public roads.

The concentration of production at the city centre was undermined by the declining cost of inter-city trucking, a development that was particularly helped by the construction and expansion of highway systems.

Similar developments arose in the USA and Europe but were slower and less pronounced on the latter continent. A major reason for these differences lies in the durability of urban capital stock in general and urban transport infrastructure in particular. This lasting impact of urban infrastructure was coupled with a slower pace of urbanisation due to: (i) a less rapid transition from an agrarian to an industrialised society in some European countries, and (ii) to the fact that European cities are much older, with historically established city centres containing a greater mixture of dwellings and businesses at the core. However, in Europe as in the US there has been a massive process of suburbanisation, which has given rise to substantial controversy as to whether or not its social cost outweighs its benefits.

Along with the evaluation of the changes in urban size and form, there are contrasting views on how urban transport policy should accommodate, contain or otherwise guide the processes of suburbanisation.

Those who are concerned about the surface growth of city areas or the decrease in population densities associate these trends with a long list of negative effects, making them difficult to evaluate. The perceived costs stem from the loss of open space, decaying historical urban structures, urban air and water pollution, traffic congestion, the loss of a sense of community, patchwork housing developments on what was once agricultural land, the separation of residential and work locations, greater public investment requirements due to spreading urban developments and, last but not least, an increasing reliance on private car use (Nechyba and Walsh, 2004).

At least part of the negative list seems to have appeared by accident or mistake, and not through attempts to reap private benefits. Strong transport policy conclusions have been drawn from negative views of the current trends of suburbanisation. The UK Urban Task Force, for example, recommended that 65% of all public expenditure for transport should be spent on projects that benefit pedestrians, cyclists and public transport users (Urban Task Force, 1999). Where urban form is concerned, it is recommended that: "*Towns and cities should be well designed to be more compact and connected, support a range of diverse uses within a sustainable urban environment which is well integrated with public transport and adaptable to change*." It is not unusual that measures to change the attitudes of transport system users are postulated: "*The renaissance will require a change of culture – through education, debate, information and participation. It is about skills, beliefs and values, not just policies* (Ibid., p. 3)."

For the US, some analysts saw an endogenous return to a lifestyle associated with dense urban developments, the advent of a "new urbanism".

Recently, some economics literature has emerged postulating a more detailed and quantitative assessment of the costs and benefits of urban sprawl, or of the costs and benefits of changing the trends in urban development, *inter alia,* by transport policy measures. The argument emphasizes the identification and quantification of the benefits of the trends towards suburbanisation, and provides a critical review of the claim that, while individuals perceive private benefits from the ongoing changes in urban structure, the social costs outweigh those benefits (Kahn, 2006; Glaeser and Kahn, 2004). Moreover, increasing efforts are being devoted to a research programme designed to determine the importance of urban form (and the system of cities) for the overall competitiveness of national economies and for long-run economic growth rates (Henderson, 2005). Productivity effects result from changing urban structures, such that a maximum of agglomeration economies materialise. These can result from exploiting increasing returns to scale in the provision of public facilities and public services as well as from increasing returns from manufacturing production. The close connection between urban and national economic development was recognised by Lucas (1988) and inspired by the development of endogenous growth models. To the extent that endogenous growth is based on knowledge spillovers and sharing between researchers and producers, and given the importance of face-to-face communication and the requirement of close spatial proximity, much of the interaction and knowledge sharing must occur at the level of individual cities.

The objective of the Round Table was to discuss these recent developments in the perspective of informing transport policies.

There is no unique way to measure urban sprawl. How it is measured is strongly influenced by whether a monocentric urban structure is perceived to be the norm or not. Close to the monocentric view of urban structure is the measure of the share of employment within a certain radius of the Central Business District (CBD) (Glaeser and Kahn, 2004).

A more comprehensive measure has been proposed by Ewing, Pendall and Chen (2005). To construct an index of urban compactness, they combine:

- residential density;
- neighbourhood mix of homes, jobs and services;
- strength of activity centres and downtowns; and
- accessibility of the street network.

This index is a more general measure of sprawl, in that it can capture the polycentric character of metropolitan areas. Based on this index, Kahn (forthcoming) presented indicators of "benefits of sprawl" for four classes of compactness for urban areas (high sprawl, sprawl, low sprawl, very low sprawl).

A first difference in consumption patterns and associated benefits concerns home ownership propensities and land consumption. Controlling for other factors which influence consumption, home ownership rates are 8.5 per cent higher in the most sprawled cities relative to the most compact city type. In compact cities, residential lots are about 40 per cent smaller than those of the median household living in a sprawled city. This does not show by how much households value such gains, as households which live in compact urban areas might have different preferences for house sizes compared to those who live in low-density settlements. However, a more compact city would lead to higher land rents, with a negative impact on the real incomes of all inhabitants.

The Round Table discussed the distributional effects of sprawl, or the distributional effects of containing sprawl by appropriate transport or other smart growth policies (e.g. Quigley and Raphael, 2005). Incumbent homeowners benefit from the increase in land rents which could result from higher intra-urban transport costs, as long as the locations of jobs and services remain fixed.

Low-income groups, with limited opportunities for wealth accumulation, tend to suffer from higher land rents. For the US, when comparing the minority/majority housing consumption differential in compact cities, it has been found that minorities who live in sprawled cities catch up in some housing consumption dimensions to majority households (Kahn, 2001, forthcoming).

COMMUTING

Much of the concern about urban sprawl has to do with an expected or observed increase in private car use and the associated increase in air pollution. This is based on the assumption that in compact cities people are likely to live closer to their downtown jobs, and that more people use public transit. It is also based on the expectation that sprawl increases congestion, leading to low private car commuting speeds, with high time losses and high costs in terms of value of time lost. As shown in two of the background papers (Kahn forthcoming, Banister forthcoming), these hypotheses cannot be confirmed in general. For the US, it was found that compared to workers in compact cities, workers in sprawled cities indeed commute over longer distances (1.8 miles further each way) but that their commute times are shorter (4.3 minutes on average), as they travel at higher speeds. The effect of this commuting pattern on air pollution is, *a priori,* ambiguous as longer distances mean more pollution for a given speed, and a higher speed may imply lower emissions per unit of distance.

A closer look at the commuting patterns in the US reveals that it may be misleading to discuss sprawl and the associated commuting patterns on the basis of the general presumption of a sprawling, monocentric structure (Anas, Arnott and Small, 1998).

A combination of the information provided by the US Neighborhood Change Database and the information on distances from the Central Business District provided by the census tracts, revealed that the share of commuters with a short commute declines over the distance 0 to 10 miles from the CBD. From the 11th mile from the CBD, the share of commuters with a short commute stops declining. An increasing share of workers with residences distant from the CBD stop commuting to the Central District. This might reflect the fact that with an expanding city size, initially through households relocating from inner city areas to outskirts, after some time, jobs follow the households, manifesting in the increasing importance of polycentric changes of urban form.

This suggestion is strongly confirmed with a closer look at US and European cases relating urban transport, and in particular commuting patterns, to settlement size, population density, the job-housing balance and mixed-use development, as well as accessibility and neighbourhood design. These four characteristics of urban areas are seen as the central control instruments of urban planners (Banister, forthcoming). The UK National Travel Survey, for example, revealed a clear correlation between settlement size and a decrease in travel distances. Looking at individual metropolitan areas, London turned out to be a special case in that commuting distances did not stop increasing when the distance of residential location from city centre increased beyond a certain threshold. For Birmingham and Manchester, the threshold distances were seven and five kilometres, respectively.

Both settlement density and the ratio of jobs to workers in a (sub-) urban region seem to have little effect on travel behaviour in general, and commuting behaviour in particular. The design of transport networks seems to have strong effects on travel patterns. The accessibility of public transit stops plays a major role in containing private car use.

Urban street design can have ambiguous effects as an instrument to reduce the demand for sprawl: while a “loops and cul-de-sac” design increases the amount of usable land, and thereby could increase density relative to a grid network (Grammenos and Tasker Brown, n.d.), the latter seems to have the advantage of increasing walking and cycling in cities (Boarnet and Crane, 1999b; Marshall, 2005).

PRODUCTIVITY AND GROWTH EFFECTS OF URBAN SPRAWL

Despite a vast literature on agglomeration effects and associated concepts of "optimal city size" as balancing economies and diseconomies, the discussion of the pros and cons of an expansion of urban areas had only little reference to this normative concept of urban form. (As an example, see Prudhomme and Lee, 1999.) One reason why economic activity agglomerates into cities is the provision of indivisible local public goods whose use is associated with transport costs. More importantly, agglomeration is due to the external benefits of production and consumption activities of firms and households. These drivers of agglomeration are, at the same time, the determinants for long-run growth rates of national economies. Consequently, urban size and urban form might strongly influence the aggregate, national growth process. Moreover, with urban form being the result of the endogenous location decisions of firms and households, the pattern of urbanisation determines the efficiency of the growth process (Black and Henderson, 1999a). This section reflects the arguments that have been raised on the link between urban form and productivity in the Round Table discussions.

External scale economies, i.e. the positive effects of the production of one firm or industry on the production of another firm or industry (Romer, 1986), or knowledge spillovers which increase the returns of private investment in education, training and research (Lucas, 1988), drive long-run increases in productivity. Early work to explain how such spillovers affected urban form simply assumed a spatial decay of the positive external effects (Fujita and Ogawa, 1982). Only recently has there been progress in providing microfoundations for such a decay (see the review in Duranton and Puga, 2004).

- A first source of city size advantages derives from the fact that the higher the level of local production, the higher will be the number of locally supplied intermediate goods. The greater the variety of intermediate goods, the greater will be the productivity of the industries using those goods. Modeling of this mechanism in the urban context assumes that increasing the congestion costs of workers commuting to the Central Business District will ultimately exhaust the benefits resulting from a greater variety of inputs (Abdel-Rahman and Fujita, 1990).

- Secondly, in an argument going back to Adam Smith (1776), the increase in the number of workers in one firm, due to an increased scale of production, allows the workers to specialise on a narrower set of tasks. The resulting productivity increase is due to workers' "learning-by-doing" effects. Moreover, the switching between tasks in production is associated with fixed switching costs, which are saved in the case of a greater specialisation. And finally, a greater specialisation on a small set of tasks allows for more technical change, as simpler tasks can be mechanised more easily (Duranton, 1998; Becker and Henderson, 2000a; Becker and Murphy, 1992). A reduction of transport costs by reducing congestion costs in transport or increasing the supply of public transport, potentially widens the market per firm and allows for a greater specialisation of the work force.

- A third positive productivity effect might result from the fact that lower urban transport costs improve the working of the labour market. A positive productivity effect is brought about by the fact that an increase in the number of firms and households trying to find a superior working relation, enhances the expected quality of a match (Helsley and Strange, 1990) and the likelihood of finding such a match (Mortensen and Pissarides, 1999; Berliant *et al.*, 2000b). The pool of interacting firms and households is limited by commuting costs or, in the longer term, by relocation costs.

– A dynamic productivity effect is expected from cities providing opportunities to enhance production-relevant knowledge. Hypotheses on the positive effects of low transport costs on the creation and dissemination of technical and organisational knowledge are based on the perception that learning is not only an individual activity but involves interaction with others, much of which is of a face-to-face nature. Cities, by bringing together large numbers of people, should therefore facilitate the production and use of technical and organisational knowledge. The smaller the intra-urban transport costs, the greater is the potential number of interacting parties.

- Knowledge diffusion is mainly considered to occur via a knowledge transfer from skilled workers to lower skilled and young workers. One mechanism, as in Jovanovic and Rob (1989), is that low-skilled workers increase their skill level by successful face-to-face interaction with skilled workers. The number of contacts between the skilled and unskilled increases with city size (Glaeser, 1999). The smaller the urban transport costs, the higher would be the number and quality of the interactions between the skilled and unskilled labour forces.

- City growth has been considered to be based on the dissemination of all workers' knowledge rather than on the transmission of knowledge from skilled to less skilled workers. The learning abilities of individual workers depend on the level of knowledge already achieved and the aggregate stock of knowledge that is available in the city. The latter provides dynamic external benefits to the workers (Lucas, 1988; Eaton and Eckstein, 1997). At least for the US, there is strong empirical evidence that the presence of educated populations in cities drives their further growth (Simon and Nardinelli, 2002; Glaeser and Saiz, 2004).

The arguments on the advantages of city size might suggest that the accommodation of an increasing city size by transport policy leads to the productivity and growth effects mentioned above. Such a conclusion is, however, in contrast to some analysis that sees population densities rather than city size as the main determinant of dynamic efficiency in production. Ciccone and Hall (1996) argue the importance of population density for productivity in a more general context, based on an empirical study. Lucas and Rossi-Hansberg (2002) also emphasize density as a driver of productivity. These arguments suggest that sprawl, a reduction of urban density, could indeed reduce agglomeration economies and therefore negatively impact on aggregate productivity. What makes the tension between the arguments asserting the importance of size and density difficult to resolve is the fact that the latter depend on the choice of the geographic area of study. Glaeser and Kahn (2004), for example, conclude that aggregate density at the metropolitan area level matters in explaining variations in per-capita income across cities, but the degree to which jobs are centralised in a Central Business District seems to be irrelevant.

Firms which are able to split management, R&D and production locations, increasingly site non-management occupations at the edge of major cities (Rossi-Hansberg, Sarte and Owens, 2005). These firms are likely to gain greatly from extensions of the city size area.

What complicates the relationship between productivity, growth and urban form further is the fact that the monocentric urban form increasingly gives way to polycentric structures. In addition, and parallel to this development, "centres" change their socio-economic function over time. As was discussed by the Round Table and argued in one of the background papers, the process of land development shares some similarities with slash-and-burn agriculture (Duranton, forthcoming). For commercial developments, economic change (sectoral decline, new technologies, etc.) typically involves leaving a vacant or under-utilised, developed site behind. Changes of urban form and structure involve some element of "creative destruction". Because real developments are highly durable, the creative destruction of production activities and firms implies a movement or re-use of company buildings and possibly a partial or complete desertion of land. The US Environmental Protection Agency, using a restrictive definition and focusing on commercial sites, estimates that there are about 450 000 brownfield sites in the US. British authorities estimate that there are 660 square kilometres of brownfield sites in England alone. Only a small part of the brownfield sites is redeveloped.

City governments or developers have to choose whether to redevelop a brownfield site or initiate new developments on a greenfield site. They face a trade-off between redeveloping a brownfield site, which may allow a better use of existing infrastructure but is maybe associated with high demolition and clean-up costs, and a greenfield development that requires new public infrastructure. From a commercial point of view, a relocation to a greenfield site may look advantageous because the costs for the required infrastructure are not, or not fully, charged to the local users, while firms often have to bear the full redevelopment costs. This allocation problem sometimes extends to communal land use and transport policy decisions, when fiscal redistribution implies that part of the infrastructure costs are borne by non-local taxpayers.

THE SOCIAL COSTS OF URBAN SPRAWL

Parallel to the progress of research on the economic benefits of the current changes in urban form, there is a continuing discussion about the social costs. The debate proceeds on distinct levels. A first level concerns the basic discussion of what should be the foundation of urban and transport policy objectives. More concretely, it tries to find an answer to the question whether individual benefits or some aggregate of individual benefits should be the only or dominant determinant of policy objectives. Often implicitly, the debate seems to evolve around the question of whether governments should supply "meritoric" goods, i.e. goods that have a social value distinct from and beyond the individual perception of their benefits. More generally, such normative arguments are related to an organic understanding of the state (Popper, 2003; Wilson, 1942). The Round Table discussion focussed on the quantitative dimension and the consequences of the social costs of sprawl. This mainly concerns the loss of farm and forest land, the consequences of urban sprawl for the transport system, and the effects of the changes of both land use and the transport system on the environment and public health (Deakin, forthcoming).

The loss of farm- and forestland

US Census data provide the opportunity to quantitatively assess the loss of open space in the form of farmland and forests due to the extension of urban space. In overall terms, the loss does not seem to be dramatic: Over the period 1974 to 2002, the total number of farmland acres in the US declined by about 8 per cent, according to the Census of 2004. Not all of the decrease was due to expanding cities but to changes in agricultural technologies, changing international competitiveness and restrictions on the provision of agricultural subsidies for some farm products. The US Department of Agriculture estimates the average annual decline to be 0.25 per cent over the 1960-2002 period.

What potentially amounts to a more substantial effect for the agricultural sector is the fact that prime farmland has been converted at two to four times the rate of less productive farm land. The loss of prime farmland is considered to be due to the competition between agrarian and urban interests in land use (USDA, 1999). The loss of forests due to urban developments is in some areas greater than forestland preserved to protect the habitat for a variety of flora and fauna, including endangered and threatened species (US Department of Agriculture, Forest Service, 2006).

These problematic trends have been mitigated to some extent by new markets in land development rights (Kahn, forthcoming): Throughout the US, municipalities are purchasing open space around their borders to guarantee that the land is not developed. The city of Boulder, Colorado has, for example, earmarked a 0.73 per cent sales tax to fund the purchase of open space around city borders to avoid it being developed. Whether and how such initiatives occur depends on the political influence of groups with an interest in new land developments and those who prefer greenbelts surrounding cities. Richer communities and jurisdictions with more home-owners seem to be more likely to initiate greenbelt initiatives (Kotchen and Powers, 2006).

Immediate costs of urban sprawl for the transport system

About 90 per cent of all person-trips in the US are made by automobile, and trucks account for more than 90 per cent of all shipments. From 1970 to 2000 the number of vehicle miles travelled has doubled and truck travel has tripled. The increasing road transport intensity, due to road transport demand growth being greater than population growth, is associated with the expectation that the public transport infrastructure cost per head is increasing. The demand for physical infrastructure is also expected to increase with the rising number of vehicle miles travelled. This is due to the relative decline of public transport use and walking as a consequence of urban sprawl.

The reduced commuting times in sprawled urban areas reported above are sometimes expected to be of a transitory nature, giving way to congestion with increasing congestion costs when scattered suburbanisation is followed by subsequent infill and development. The relatively high commuting speeds are then no longer sustainable (Cervero, 1986; Landis and Reilly, 2003).

Low-density development, and the emergence of a polycentric urban structure, makes it difficult and costly to provide bus, light rail or metro services. The increased private car use required by these urban forms is sometimes held to lead to greater resource demands for transport than a transport system with a higher share of public transport and a different settlement pattern.

Environmental costs of the transport consequences of urban sprawl

There is no disagreement that changes in urban form, which reduce the compactness of cities and lower the settlement density, increase the vehicle-miles traveled by individual households and reduce the share of public transport usage. Both effects contribute to current changes of urban form being associated with higher environmental costs for transport. Greenhouse gas emissions from transport are a function of fuel use. In the US, transport is currently responsible for 32 per cent of total carbon emissions. Moreover, its emissions from transport increase by 1 to 2 per cent annually.

Air pollution more generally remains a public health concern. To some extent, this is due to inadequate responses to more restrictive air quality regulations. The full health consequences of air toxins and fine particles have not always led to the required technical standards for transport equipment.

The relationship between urban form and emissions is complicated by the fact that emissions are not a simple product of speed. Stop-and-go traffic, which might result from congestion in compact cities, is more polluting than steady-flow traffic. On the other hand, very high speeds, which might be associated with sprawl and metropolitan highways, also produce very high emission levels.

While not necessarily providing an argument against the environmental concerns relating to low-density settlements, vehicle emissions regulation has been able to offset increased vehicle mileage due to changing settlement patterns. The Los Angeles Basin, for example, suffers from the highest levels of air pollution in the US, mainly caused by vehicle emissions. The area is, at the same time, a prime example of low-density, car-dependent urban development (Giuliano and Small, 1991). But ambient ozone, a leading indicator of smog, declined by 55% between 1980 and 2002, from 0.21 to 0.095 parts per million on average for the country's nine monitoring stations. This decline occurred despite an increase in population of 29 per cent in the same period of time and a 70 per cent increase in total automobile mileage (Kahn, forthcoming). Due to developments in vehicle technology, population growth in low-density areas was not necessarily associated with higher air pollution. Kahn found a negative correlation between country population growth and increased ambient air pollution for California over the years 1997-2002.

Current research shows that relationships between urban form, infrastructure design and travel behaviour are still not fully understood. Much of the research into the relationship between the transport sector and urban form has focussed on physical effects. It is even more demanding to identify the valuations of external costs and benefits of different urban forms. Only a full evaluation of external effects would allow to draw final conclusions on whether current changes of urban form provide net benefits and how these should be maximised by transport policy action.

CONCLUSION

The Round Table discussed recent research that throws light on the benefits derived and costs incurred through changes of urban form. The development of city sizes and structures is driven by the design of the transport system, and/or confronts transport policy with demands to accommodate or contain ongoing changes of land use.

On the benefit side, the discussion identified two main benefits to be drawn from current trends towards suburbanisation, which amount to an increase in city size, and a decreasing housing and population density in urban areas:

- The decline in housing density has clearly increased the number of vehicle miles travelled. However, beyond a certain threshold distance from traditional city centres, commuting times decrease. While passenger transport became more infrastructure-intensive, travel, and in particular commuting times, decreased. Higher infrastructure investment has led to time savings, owing to reduced congestion. The effect of these changes on fuel consumption is ambiguous: less congestion might lead to lower fuel consumption if higher speeds remain within an intermediate range.

- Households living in low-density settlements, with relatively small land rents, have higher rates of home-ownership and consume more residential land. This has particularly benefited low-income households.

Agglomeration economies are central to the argument that an increase of city size increases the productivity of goods and services. Decreasing transport costs are considered to be instrumental for the spatial extension of the mechanisms leading to agglomeration economies:

- An increase in city size might increase the availability of specialised inputs. This in turn increases the productivity of final goods production.

- An increasing city size driven by lower transport costs might allow a greater specialisation of the work force, leading to productivity effects associated with "learning-by-doing".

- Lower passenger travel costs within metropolitan areas may increase the performance of the labour market. A more highly mobile workforce is expected to increase the probability and quality matches between employers and workers.

Dynamic agglomeration economies have recently received particular attention. The larger the cities, and the easier the interaction between skilled and unskilled workers or knowledge-producing agents, the higher is the rate of knowledge diffusion, and the higher will be the rate of knowledge production. Both determine the long-run growth of urban as well as national economies. To the extent that the ease of interaction between individuals who transmit or jointly produce knowledge depends on density, urban sprawl might negatively affect growth. This is strongly influenced by firms deciding to separate management, R&D and production locations. The more companies can split production from research and management, the more they will benefit from increasing city sizes.

Intensive research efforts have led to a great awareness of the costs of urban sprawl. Many effects are, however, context specific. A major part of the research focuses on the physical consequences of urban design and the design of transport systems.

– A first social cost of the current trend toward urban development is seen in the loss of farm- and forestland. While the annual percentage decline of farmland is rather small, some concern exists about the loss being concentrated on prime farmland.

– The immediate, transport-related costs are considered to be high, and to be due to the fact that infrastructure costs are not internalised by users of the transport system. A similar argument is made for congestion costs. The reduction of time losses due to congestion is expected to be a temporary phenomenon, which will disappear with the filling in of vacant land.

– Environmental costs and air pollution, which are due to the augmentation of vehicle-kilometres and the decline of public transit patronage, remain major concerns of critics of increasing city size. This criticism is maintained, despite the strong emission reductions that have been observed in metropolitan areas over the last decade. Rapid developments in car technology, often induced by more restrictive regulations, have led to a reduction of emissions despite the increase in vehicle miles travelled associated with urban sprawl.

REFERENCES

Abdel-Rahman, H.M. and M. Fujita (1990), Product variety, Marhallian externalities, and city sizes, *Journal of Regional Science*, 30: 165-183.

Anas, A., R. Arnott and K. Small (1998), Urban spatial structure, *Journal of Economic Literature*, 36: 1426-64.

Banister, D. (forthcoming), Cities, urban form and transport, in: OECD/ECMT Transport Research Centre (ed.), *Transport, Urban Form and Economic Growth, Round Table 137*, Paris.

Barrett, P. (1983), *The Automobile and Urban Transit: The Formation of Public Policy in Chicago, 1900-1930*, Philadelphia.

Becker, G.S. and K.M. Murphy (1992), The division of labor, coordination costs, and knowledge, *Quarterly Journal of Economics*, 107: 1137-1160.

Becker, R. and J.V. Henderson (2000), Intra-industry specialization and urban development, in: J.-M. Huriot and J.-J. Thisse (ed.), *Economies of Cities: Theoretical Perspectives*, Cambridge, Mass.

Berliant, M., R.R. Reed and P. Wang (2000), Knowledge exchange, matching, and agglomeration. Federal Reserve Bank of Minneapolis Discussion Paper 135, Minneapolis.

Black, D. and V. Henderson (1999), A theory of urban growth, *Journal of Political Economy*, 107: 252-84.

Boarnet, M.G. and R. Crane (1999), *Travel by Design: The Influence of Urban Form on Travel,* New York.

Cervero, R. (1986), *Suburban Gridlock*, New Brunswick, N.J.

Ciccone, A. and R.E. Hall (1996), Productivity and the density of economic activity, *American Economic Review*, 86: 54-70.

Deakin, E. (forthcoming), The social cost of urban sprawl, in: OECD/ECMT Transport Research Centre (ed.), *Transport, Urban Form and Economic Growth, Round Table 137*, Paris.

Duranton, G. (1998), Labor specialization, transport costs, and city size, *Journal of Regional Science*, 38: 553-573.

— (forthcoming), The insatiable demand for land: Urban change and land (re-)development, in: OECD/ECMT Transport Research Centre (ed.), *Transport, Urban Form and Economic Growth, Round Table 137*, Paris.

Duranton, G. and D. Puga (2004), Micro-foundations of urban agglomeration economies, in: J.V. Henderson and J.-F. Thisse (ed.), *Cities and Geography. Handbook of Regional and Urban Economics,* Vol. 4, Amsterdam.

Eaton, J. and Z. Eckstein (1997), Cities and growth: Theory and evidence from France and Japan, *Regional Science and Urban Economics,* 27: 443-474.

Ewing, R., R. Pendall and D. Chen (2005), Measuring Sprawl and its Impact,

www.smartgrowthamerica.org

Fujita, M. and H. Ogawa (1982), Multiple equilibria and structural transition of non-monocentric urban configurations, *Regional Science and Urban Economics*, 12: 161-196.

Giuliano, G. and K.A. Small (1991), Subcenters in the Los Angeles Region, *Regional Science and Urban Economics*, 21: 163-82.

Glaeser, E. and M. Kahn (2004), Sprawl and urban growth, in: V.L. Henderson and J.-F. Thisse (ed.), *Cities and Geography. Handbook of Regional and Urban Economics,* Vol. 4, Amsterdam.

Glaeser, E.L. (1999), Learning in cities, *Journal of Urban Economics*, 46: 254-277.

Glaeser, E.L. and A. Saiz (2004), The rise of the skilled city, *Brookings Wharton Papers on Urban Affairs,* 5: 47-94.

Grammenos, F. and J. Tasker Brown (n.d.), Residential street pattern design for healthy, livable communities, New Urban Agenda, www.greenroofs.ca/nua/ip/ip02.htm

Helsley, R.W. and W.C. Strange (1990), Matching and agglomeration economies in a system of cities, *Regional Science and Urban Economics*, 20: 189-212.

Henderson, J.V. (2005), Urbanization and growth, in: P. Aghion and S.N. Durlauf (ed.), *Handbook of Economic Growth,* Vol. 1b, Amsterdam.

Jovanovic, B. and R. Rob (1989), The growth and diffusion of knowledge, *Review of Economic Studies*, 56: 569-582.

Kahn, M. (2001), Has sprawl reduced the black/white housing consumption gap?, *Housing Policy Debate*, 12: 77-86.

— (2006), *Green Cities: Urban Growth and the Environment,* Washington, DC.

— (forthcoming), Quality of life and productivity in sprawled versus compact US cities, in: OECD/ECMT Transport Research Centre (ed.), *Transport, Urban Form and Economic Growth, Round Table 137*, Paris.

Kotchen, M.J. and S. Powers (2006), Explaining the appearance and success of voter referenda for open space conservation, *Journal of Environmental Economics and Management,* 52: 373-390.

Landis, J.D. and M. Reilly (2003), How we will grow: baseline projections of the growth of California's urban footprint through the year 2100, Institute of Urban and Regional Development, University of California at Berkeley, Berkeley.

Lucas, R.E. (1988), On the mechanics of economic development, *Journal of Monetary Economics*, 22: 3-42.

Lucas, R.E. and E. Rossi-Hansberg (2002), On the internal structure of cities, *Econometrica*, 70: 1445-76.

Marshall, S. (2005), *Streets and Patterns*, London.

Mortensen, D.T. and C.A. Pissarides (1999), New developments in models of search in the labor market, in: O. Ashenfelter and D. Card (ed.), *Handbook of Labor Economics,* Vol. 3, Amsterdam.

Nechyba, T.J. and R.P. Walsh (2004), Urban Sprawl, *Journal of Economic Perspectives*, 18: 177-200.

Popper, K. (2003), *The Open Society and its Enemies, Vol. I. The Spell of Plato,* New York.

Prudhomme, R. and C.-W. Lee (1999), Size, sprawl, speed and the efficiency of cities, *Urban Studies*, 36: 1849-1858.

Quigley, J.M. and S. Raphael (2005), Regulation and the high cost of housing in California, *American Economic Review*, 9: 323-328.

Romer, P. (1986), Increasing returns and long-run growth, *Journal of Political Economy*, 94: 1002-1037.

Rossi-Hansberg, E., P.-D. Sarte and R. Owens III (2005), Firm fragmentation and urban patterns, NBER Working Paper No. 11839, Cambridge, Mass.

Simon, C.J. and C. Nardinelli (2002), Human capital and the rise of American cities, *Regional Science and Urban Economics*, 32: 59-96.

Smith, A. (1776), *An Inquiry into the Nature and Causes of the Wealth of Nations*, London.

US Department of Agriculture (1999), Extent of Federal Influence on “Urban Sprawl” Is Unclear, Washington, DC.

US Department of Agriculture, F.S. (2006), The Highlands Area of Connecticut, New Jersey, New York and Pennsylvania, Washington, DC.

Urban Task Force (1999), *Towards an Urban Renaissance, The Report of the Urban Task Force*, Chaired by Lord Rogers of Riverside, London.

Wilson, F.G. (1942), The revival of organic theory, *American Political Science Review*, 36: 454-459.

– – – – –