VTI meddelande 963A · 2004

Measuring the marginal cost of road use
An international survey

Nils Bruzelius

Swedish National Road and Transport Research Institute

VTI meddelande 963A · 2004

Measuring the marginal cost of road use

An international survey

Nils Bruzelius

Swedish National Road and Transport Research Institute

Publisher:

Swedish National Road and Transport Research Institute

SE-581 95 Linköping Sweden

Publication:
VTI meddelande 963A

Published: 2004

Project code: 91000

Project:
Implementing marginal cost based pricing in the transport sector

Author:
Nils Bruzelius

Sponsor:
Vinnova
Vägverket (Swedish National Road Administration)
Banverket

Title:
Measuring the marginal cost of road use – an international survey

Abstract (background, aims, methods, results) max 200 words:

The Swedish National Road and Transport Research Institute (VTI) has carried out this survey, which is expected to generate results of use for implementation of marginal cost based pricing of transport in terms of the principles laid down in the white paper on a common transport policy presented by the EU Commission in September 2001. The purpose of the survey is to review alternative methods that may be used for estimating marginal infrastructure costs.

The survey identifies four different approaches to the measurement of the marginal cost of road use and evaluates empirical studies based on these approaches. One approach is referred to as the econometric approach. There are few examples of this approach, which is explained by that it is difficult to generate the required data. The value of the approach is also limited by that it relies on historical data.

Two of the other methods differentiated between are often based on similar assumptions. It is argued that both what is called the indirect approach and the full cost allocation approach basically rely on the properties of Newbery's 'fundamental theorem'.

A fourth approach identified in the Survey is to use the models (PMSs) used by road authorities in the planning of road works. The idea is not new, but has rarely been applied. It is demonstrated in the survey how the approach may be employed for a specific road by way of the HDM 4 model, developed by the World Bank. There are several advantages of using a PMS, including that marginal costs will be derived with the same methodology that is used to rank investments, determine maintenance priorities and to prepare a road authority's budget.

ISSN:	Language:	No. of pages:
0347-6049	English	54 + Appendix

Utgivare:

Väg- och transport-forskningsinstitutet

581 95 Linköping

Publikation:
VTI meddelande 963A

Utgivningsår: 2004

Projektnummer: 91000

Projektnamn:
Implementering av marginalkostnadsprissättning i transportsektorn

Författare:
Nils Bruzelius

Uppdragsgivare:
Vinnova
Vägverket
Banverket

Titel:
Att beräkna marginalkostnaden av vägutnyttjande – en internationell översikt

Referat (bakgrund, syfte, metod, resultat) max 200 ord:

Statens väg- och transportforskningsinstitut (VTI) har tagit fram denna översikt för att skaffa ytterligare insikter som är användbara vid en implementering av marginalkostnadsbaserad prissättning av transporter enligt de riktlinjer som lades fast i den vitbok om en gemensam transportpolitik som presenterades av den Europeiska Kommissionen i september 2003. Syftet med denna översikt är att studera alternativa metoder som kan användas för att skatta marginella infrastrukturkostnader.

I översikten identifieras fyra olika ansatser för att mäta marginalkostnaden av vägutnyttjande och utvärderas empiriska studier baserade på dessa ansatser. En ansats benämns den ekonometriska ansatsen. Det finns få exempel på denna ansats, vilket förklaras av att det är svårt att generera den nödvändiga informationen. Värdet av denna ansats begränsas också av att den är baserad på historiska data.

Två av de andra metoderna bygger ofta på liknande antaganden. I översikten hävdas att såväl den så kallade indirekta ansatsen som kostnadsallokeringsansatsen egentligen baseras på de egenskaper som ligger till grund för Newberys 'fundamentala teorem'.

En fjärde ansats, som identifieras i översikten, är att nyttja de modeller (PMS) som används av väghållare vid planeringen av väghållningsåtgärder. Idén är inte ny, men den har sällan använts. I översikten visas hur denna ansats kan tillämpas på en specifik vägsträcka med hjälp av Världsbankens HDM 4 modell. Det finns flera fördelar med att använda en PMS, inklusive att marginalkostnaderna beräknas med samma metod som används för att rangordna investeringar, prioritera underhållsinsatser och bereda väghållarens budget.

ISSN:	Språk:	Antal sidor:
0347-6049	Engelska	54 + Bilaga

Preface

Economic principles have been a pillar of Swedish transport policy for decades. The policy prescribes that decisions on infrastructure investments shall be based on cost benefit analysis (CBA) and that pricing of infrastructure use shall be based on socioeconomic marginal costs. The pricing rule has recently also been included into European transport policy where marginal cost pricing is seen as key to the promotion of an open transport market. The last few years have also seen the technology for more advanced pricing methods rapidly improving, and we will in the near future therefore see the implementation of systems that allow advanced pricing principles to be realised. Considerable research has focused on CBA principles including research on the value of time and the value of statistical life. However, relatively limited research has previous been conducted on the pricing principles of the policy.

With this background, the Swedish National Road and Transport Research Institute (VTI) carries out a three year research project – 'Implementing marginal cost based pricing in the transport sector' – financed by Vinnova, Banverket and Vägverket[1], with the aim to improve the knowledge about the external marginal cost of the transport sector. The project covers all modes of transport and several cost categories, such as infrastructure cost, accident cost and environmental cost.

One of the important components of a pricing policy for roads is the marginal cost of road use. This report, written by Dr. Nils Bruzelius, summarises the state-of-the-art of the marginal cost of road use and is an important input into our further research in this area. It is surprising to see that so many questions remain to be answered about the marginal cost of road use, an issue which has been discussed for decades and is important to road authorities all over the world. Dr. Bruzelius makes in this report a useful overview over the state-of-the-art and a classification of different approaches that has been and can be used to estimate the marginal cost of road use. Our further research in this area will hopefully answer some of the questions that are raised.

Stockholm, February 2004

Gunnar Lindberg
Project Leader

[1] Swedish Agency for Innovation Systems, Swedish National Rail Administration and Swedish National Road Administration

Table of Contents Page

Measuring the marginal cost of road use – an international survey

by Nils Bruzelius
Nils Bruzelius AB
Mätaregränden 6, SE-226 47 Lund, Sweden

Executive Summary

The National Swedish Road and Transport Research Institute (VTI) has carried out this survey, which is expected to generate results of use for implementation of marginal cost based pricing of transport in terms of the principles laid down in the white paper on a common transport policy presented by the EU Commission in September 2001. The purpose of this Survey is to review alternative methods that may be used for estimating marginal infrastructure costs, as well as results obtained with these methods.

The short-run marginal infrastructure cost (or marginal cost of road use) associated with an additional vehicle on a road comprises three components. These are:

The increase in the cost inflicted on other vehicles as a consequence of the additional vehicle. This cost, referred to as the **road damage externality**, reflects that the vehicle will cause some wear and damage to the road, and that this deterioration will result in increased costs to subsequent vehicles in the form of (i) increased cost for operating the vehicles, (ii) lower speeds, resulting in higher time costs, and/or (iii) less comfort when driving the vehicle.

The fact that additional **wear** of the road will lead to the road authority taking action to remedy the road wear at an earlier data than would have been the case without the additional vehicle. It is assumed that the road authority monitors road wear and takes action when the effects of road wear, e.g. in the form of reduced friction, results in a condition worse than a certain critical level. For there to be a marginal cost associated with road wear, it has to be assumed that (routine) maintenance actions with respect to road wear are condition responsive (the marginal cost of road wear).

The fact that additional **deformation** (or **damage**) of the road will lead to the road authority taking action to remedy the damage at an earlier date than would have been the case without the additional vehicle. It is assumed that the road authority monitors road deformation and takes action when the effects of damage, e.g. in the form of increased IRI value, reaches a certain critical level. It is thus assumed that (periodic) maintenance actions with respect to road damage is condition responsive (the marginal cost of road damage).

Four different approaches to the measurement of the marginal cost of road use may be identified. The first is referred to as the direct method. The direct method focuses on actually determining the marginal cost of road use without necessarily making a distinction between the three components. It relies on the Pavement Management Systems (PMS) used by road authorities in order to plan road maintenance and investment actions.

The second method is referred to as the indirect method. It is based on the 'fundamental theorem' formulated by David Newbery. The characteristic of this approach is that it is based on the assumption that the road damage externality cost may be ignored. It focuses on the consequences of road damage and periodic

maintenance in the form of overlays. Limited work has been made to extend this approach to apply to road wear and condition responsive routine maintenance, and reconstruction of a road. There exists, however, one frequently used approach to the measurement of the marginal cost of road use which could be seen as an extension of the Newbery approach, also considering wear and condition responsive routine road maintenance. This is the RUC30 model developed by the World Bank.

The majority of road costing exercises to be found have no clear foundation in the marginal cost approach although they contain results which could be seen to be, and often are, of relevance to the estimation of marginal costs. These other cost studies, frequently referred to as cost allocation studies, essentially have a different perspective in that they focus on equity, and variable and fixed costs of road works (i.e. costs of road works which are viewed as being a function of traffic, measured in terms of one or several of the measures of road use (see below under 2), and not being a function of traffic) as opposed to marginal and non-marginal costs associated with an additional vehicle. In this Survey these cost allocation studies are also referred to as belonging to the club approach as estimates from them can essentially be seen as trying to answer the question: Assume that all users of a road belong to a club, and that they have to agree on a system for how to recover the costs of road works by way of user charges. What would then be the characteristics of such charges?

The three previous approaches are based on estimations essentially making use of various types of unit costs. The fourth approach also relies on unit cost of sorts, but marginal costs are determined first after having estimated a cost function by way of econometric techniques. This approach builds on conventional micro-economic production theory.

The Survey leads to the following conclusions:

1. Marginal infrastructure costs are heavily influenced by the damaging effects in the form of various types of distresses caused principally by the loading of traffic. Loading is measured by way of Equivalent Standard Axle Loads (ESALs), and normally the fourth power law is made use of. This law, which implies that an axle with a load twice that of an axle which gives rise to a damaging impact of one ESAL, results in a load which is 16 times higher, is commonly used in calculations of marginal costs, but its relevance is somewhat in question. This is an important issue in that the size of the power coefficient may significantly affect the ratio of the marginal infrastructure cost for heavy and light vehicles. However, this is a common issue to all approaches used today to measure marginal infrastructure costs.

2. In the ongoing EC-sponsored UNITE project for estimating marginal costs in transport, reference is made to econometric approaches in order to measure marginal infrastructure costs, but there are few applications of this approach. The reason is likely to be that it is difficult to generate the required data – longitudinal data over a long period of time for 'segments' of roads will likely be required – and it seems unlikely that road authorities will have such data in their management information systems (MIS). Also, it will be difficult to disentangle the effects on maintenance of various measures of road use (vehicle-km, gross vehicle mass-km, ESAL-km, etc.),

as they tend to vary in the same way. Besides, the econometric approach normally (albeit not necessarily) makes use of historical data. Presumably, estimates of marginal costs should be based on current or expected future cost data, as marginal cost based pricing is supposed to guide resource use today and in the future. Mention should also be made of that the econometric approach does not account for costs for road damage externalities.

3. A distinction is sometimes made in the literature between the cost allocation and marginal cost approaches when determining costs related to road use, seemingly suggesting that the two approaches are incompatible. The Survey argues that the distinction is not necessarily substantive. Cost allocation exercises are generally based on the notion of equity; a theoretical basis for equity is the club approach, one central aspect of which is that a marginal cost should be borne by the user giving rise to this cost. Since cost allocation exercises normally are based on the club approach, they should be of interest to a review of estimates of marginal infrastructure costs.

4. Indeed, it is argued that both what is called the indirect approach and the full cost allocation approach (when based on the club approach) basically rely on the properties of Newbery's 'fundamental theorem'. It is understood that this theorem allows for ignoring the road damage externality effect, and for deriving marginal costs from estimates of average costs. It is also understood that Newbery's theorem is primarily of importance to overlays (the cost of which makes up a very large – and apparently growing – share of maintenance costs), but that it may be applied also to other maintenance costs which arise as a consequence of road use, e.g. certain routine maintenance costs, and presumably also reconstruction costs. It is noted that this possible extension of the Newbery approach apparently has never been developed in full.

5. This notwithstanding, there are other issues related to the Newbery approach, as different researchers and analysts use different methodologies, which are related but result in different conclusions. Newbery's conclusion is that marginal costs with regard to overlays may be calculated as the average cost per ESAL corrected – and reduced – to account for the weathering effect (i.e. the wear and damage caused by weather and time). The marginal cost will therefore be lower than the average cost, and optimal road user charges will therefore not recover road maintenance costs. In a study by Kenneth Small et al., a similar modelling approach as Newbery’s is used, but it is found that the marginal cost of an overlay with respect to an ESAL is higher than the average cost. Gunnar Lindberg in a Swedish study uses a different – more general – relationship than Newbery and Small et al. between the measure of a road's condition and road use measured in terms of ESALs. In terms of his findings, therefore, the average cost has to be corrected by a further term, the 'deterioration elasticity'. These differences in modelling approach are not trivial when it comes to the calculation of marginal costs. They likely have ramification also with respect to other types of maintenance and reconstruction costs, although no analysis thereof is available.

6. An important assumption underlying Newbery's 'fundamental theorem' is that maintenance actions are rule-based or condition responsive. Normally it is assumed that once roughness exceeds a certain value then the road will be overlaid. Such principles are also followed by road authorities, albeit perhaps more in theory and planning than in reality. This raises the question: If the planning framework is assumed to be relevant, and if, in addition, prices should be tailored to marginal costs so that they in effect influence present and future road use, then why not use the planning models (PMSs) used by road authorities to determine marginal infrastructure costs? The idea is not new, but has rarely (if ever?) been applied. It has been demonstrated here that it may be used, as illustrated with the application of the HDM 4 model, developed by the World Bank, for a specific road.

7. The implications of using a PMS (such as HDM 4) are the following. It has to be recognised that these models embody certain assumptions reflecting the effect of weathering and the nature of the relationship between road use and change in road condition. It is understood that with respect to overlays, the HDM 4 model is close to the assumptions used in the Newbery approach (with implications as concerns weathering and the deterioration elasticity). On the other hand, using a PMS obviates the need for relying on a partial approach as characterises the indirect method as well as the econometric approach. As demonstrated by calculations presented in this Survey, the more comprehensive approach to wear and damage reflected in a PMS such as HDM 4 could well lead to the conclusion that the road damage externality component ought not to be ignored. There are two further aspects to be noted. Firstly by using a PMS approach the marginal cost associated with a particular type of vehicle is calculated directly; it is not necessary to identify separately the marginal cost for each one of the road use variables, (assuming that ESALs do not capture all effects of road use). Secondly, it directly takes into account all types of maintenance actions, i.e. routine, periodic and reconstruction, etc., resulting from road use. Other advantages are that the approach relies on unit costs for various inputs into different types of works activities, and can be tailored – by calibration – to reflect the conditions in a particular country and part of a country, including the particular nature of roads as concerns pavement, base and sub-base. In addition, by using a PMS such as the HDM 4 it will be possible to determine marginal costs at different levels of aggregation, e.g. either for a specific road or as an average for a road network. Newbery's approach can only be applied for a road network.

8. A number of estimates of marginal costs, expressed sometimes in terms of ESALs, sometimes per km for different types of vehicles, are presented in this Survey. No effort is being made to compare the results in view of the differences in methodologies used, the nature and quality of the data employed, and that many empirical results refer to situations which may not at all be relevant to the conditions where marginal cost based pricing is to be applied. This Survey demonstrates that it is possible to obtain estimates of marginal road use costs, although the quality will vary;

however before making use of such estimates, the specifications that should be applied with respect to estimates of marginal infrastructure costs must be formulated.

Att beräkna marginalkostnader av vägutnyttjande – en internationell översikt

av Nils Bruzelius
Nils Bruzelius AB
Mätaregränden 6, SE-226 47 Lund, Sweden

Sammanfattning

Statens väg- och transportforskningsinstitut (VTI) har tagit fram denna översikt för att skaffa ytterligare insikter som är användbara vid en implementering av marginalkostnadsbaserad prissättning av transporter enligt de riktlinjer som lagts fast i den vitbok om en gemensam transportpolitik som presenterades av den Europeiska Kommissionen i september 2003. Syftet med denna översikt är att studera alternativa metoder som kan användas för att skatta marginella infrastrukturkostnader.

Den kortsiktiga marginalkostnaden (eller marginalkostnaden av vägutnyttjande) förknippad med ytterligare ett fordon på en väg består av tre komponenter. Dessa är:

Ökningen i kostnader för andra fordon till följd av det tillkommande fordonet. Denna kostnad, här kallad **trafikantexternalitet**, speglar att ett fordon ger upphov till slitage och skador på en väg och att denna nedbrytning medför ökade kostnader för efterföljande fordon i form av (i) ökade fordonskostnader, (ii) lägre hastighet, medförande högre tidskostnad och/eller (iii) sämre komfort.

Det faktum att ytterligare **vägslitage** medför att väghållaren beslutar sig för att reparera slitaget tidigare än vad som hade blivit fallet utan det ytterligare fordonet. Det antas att väghållaren mäter vägslitaget och vidtar åtgärder när konsekvenserna av slitaget, t.ex. i form av minskad friktion, resulterar i att tillståndet understiger en viss kritisk nivå. För att en marginalkostnad ska uppkomma förutsätts att (löpande) underhållsåtgärder till följd av vägslitage är betingade av tillståndet (marginalkostnaden av vägslitage).

Det faktum att ytterligare **deformation** av (eller **skador** på) vägen leder till att väghållaren beslutar sig för att åtgärda skadorna tidigare än vad som hade varit fallet utan det ytterligare fordonet. Det antas att väghållaren bevakar vägdeformationen och vidtar åtgärder när konsekvenserna av skadorna, t.ex. i form av ökat IRI värde, når en viss kritisk nivå. Det antas m.a.o. också att (periodiska) underhållsåtgärder med anledning av vägdeformation är betingade av tillståndet (marginalkostnaden av vägskador).

Fyra olika ansatser för att mäta marginalkostnaden av vägutnyttjande kan identifieras. Den första benämns här den direkta metoden. Med denna metod kan marginalkostnaden av vägutnyttjande beräknas utan att de tre komponenterna nödvändigtvis särskiljs. Den förutsätter tillgång till en PMS (’pavement management system’), dvs. den typ av modeller som nyttjas av väghållare för att planera underhålls- och investeringsåtgärder.

Den andra metoden benämns den indirekta metoden. Den baseras på det ’fundamentala teorem’ som formulerats av David Newbery. Det utmärkande för ansatsen är att den bygger på antagandet att trafikantexternaliteten kan ignoreras. Den fokuserar därigenom på konsekvenserna av vägdeformation och periodiskt underhåll bestående av ny beläggning. Få försök har gjorts för att även tillämpa

denna ansats på slitage och tillståndsbetingat löpande underhåll och ombyggnad av väg. Det finns emellertid en ofta använd ansats att mäta marginalkostnaden av vägutnyttjande vilken kan ses som en utvidgning av Newberys ansats genom att också beakta slitage och tillståndsbetingat löpande underhåll. Denna modell benämns RUC30 och har utvecklats av Världsbanken.

De flesta tillgängliga vägkostnadsstudier är inte tydligt förankrade i en marginalkostnadsansats även om de innehåller resultat som kan vara, och ofta är, av relevans för skattning av marginalkostnader. Dessa övriga studier, som ofta kallas kostnadsallokeringsstudier, har egentligen en annan utgångspunkt då de fokuserar på rättvisa och rörliga och fasta kostnader av vägunderhåll (dvs. kostnader av vägunderhåll som antas vara en funktion av trafikvolymen, mätta enligt en eller flera av de olika måtten på vägutnyttjande (se nedan under pkt. 2), och kostnader som inte är en funktion av trafik) i motsats till marginell och icke-marginell kostnad till följd av ytterligare ett fordon. I denna översikt benämns dessa kostnadsallokeringsstudier även för klubbansatsen eftersom de i allt väsentligt kan ses som ett försök att besvara frågan: Antag att alla nyttjare av en viss väg tillhör en klubb och att de därför måste komma överens om ett system för att täcka kostnaderna för vägunderhåll med hjälp av användaravgifter. Vilka egenskaper kommer att utmärka ett sådant system för vägavgifter?

Två av ansatserna är baserade på estimat vilka härletts med olika former av enhetskostnader. I den fjärde ansatsen används också enhetskostnader, men marginalkostnader bestäms först efter det att en kostnadsfunktion har estimerats med hjälp av ekonometriska metoder. Denna ansats bygger på traditionell mikroekonomisk produktionsteori.

Denna översikt leder till följande slutsatser:

1. De marginella infrastrukturkostnaderna är starkt beroende av de skador i form av olika typer av utmattning som uppkommer i huvudsak till följd av trafikbelastningen. Belastningen mäts i antal standardaxlar (ESAL), och normalt används fjärdepotensregeln. Denna regel, som implicerar att en axel med dubbla vikten jämfört med en som ger upphov till en ESAL, medför en belastning som är 16 gånger högre, används vanligen i beräkningar av marginalkostnader, men dess relevans har ifrågasatts. Detta är en viktig fråga eftersom storleken på potenskoefficienten har stor inverkan på fördelningen av marginalkostnaden mellan tunga och lätta fordon. Detta är dock en frågeställning som är gemensam för alla ansatser som används idag för att mäta den marginella infrastrukturkostnaden.

2. I det pågående UNITE projektet som finansieras av EU och som syftar till att estimera marginalkostnader av transporter, utpekas den ekonometriska ansatsen som en möjlig metod för att mäta marginella infrastrukturkostnaden, men det finns få tillämpningar av denna ansats. Orsaken torde vara att det är svårt att generera nödvändiga indata – longitudinell data över en lång tidsperiod för sektioner av vägar skulle antagligen behövas – och det är inte heller troligt att väghållare har sådana data i sina informationssystem (VDB). Dessutom, kan det vara svårt att särskilja effekterna på underhåll av olika former av vägutnyttjande (fordonskm, bruttotonkm, ESAL-km etc.), eftersom dessa variabler tenderar att samvariera. Ett annat problem är att den ekonometriska ansatsen vanligen förutsätter tillgång till historiska data. Estimat av marginalkostnader bör sannolikt baseras på

nuvarande eller förväntade framtida kostnader, eftersom marginalkostnadsbaserad prissättning förväntas styra resursanvändningen idag och i framtiden. Den ekonometriska ansatsen beaktar ej heller trafikantexternaliteten.

3. I litteraturen görs ibland en åtskillnad mellan kostnadsallokeringsansatsen och marginalkostnadsansatsen, i samband med behandling av frågor som avser kostnader relaterade till vägutnyttjande, något som skulle kunna tolkas som att de två ansatserna inte är förenliga. I denna översikt hävdas att denna distinktion inte nödvändigtvis är relevant. Kostnadsallokeringsstudier tar vanligen sin utgångspunkt i ett rättvisekoncept; en teoretisk grund för rättvisa finns i klubbansatsen i vilken en central aspekt är att marginalkostnaden ska bäras av den användare som ger upphov till denna kostnad. Eftersom kostnadsallokeringsstudier ofta använder sig av klubbansatsen borde de vara av intresse i en översikt av skattningar av marginella infrastrukturkostnader.

4. I översikten hävdas att såväl vad som kallas den indirekta ansatsen som kostnadsallokeringsansatsen (när den baseras på klubbansatsen) i huvudsak bygger på egenskaperna hos Newberys ’fundamentala teorem’. Detta teorems innebörd är att trafikantexternaliteten kan ignoreras och att marginalkostnaden kan härledas från skattningar av genomsnittskostnaden. Newberys teorem är primärt av intresse för beläggningar (vars kostnader utgör en mycket stor del – och tydligen växande andel – av underhållskostnaderna), men det kan också appliceras på andra underhållskostnader som uppkommer som en följd av vägtrafik, t.ex. vissa löpande underhållskostnader och antagligen också kostnader för vägombyggnad. Det skall nämnas att denna möjliga utvidgning av Newberys ansats inte tycks ha analyserats.

5. Det finns också andra aspekter på Newberys ansats och som fått till konsekvens att olika forskare använder sig av olika varianter av den. Newberys slutsats är att marginalkostnaden med avseende på beläggningar kan beräknas som genomsnittskostnaden per ESAL justerad – och reducerad – för att ta hänsyn till klimateffekten (dvs. att slitage och skador orsakas av klimatet). Den marginella kostnaden kommer därför att vara lägre än den genomsnittliga kostnaden, vilket innebär att optimala vägavgifter inte kommer att generera intäkter som täcker underhållskostnaderna. I en studie av Kenneth Small et.al., i vilken används en modell som liknar Newberys, kommer man fram till att marginalkostnaden av en beläggning med avseende på ESAL är högre än genomsnittskostnaden. Gunnar Lindberg använder i en svensk studie ett annorlunda – mer generellt – samband än Newbery och Small et.al. mellan vägytans tillstånd och vägutnyttjande mätt i ESAL. Lindbergs slutsats är att genomsnittskostnaden skall justeras med ytterligare en term, en ’nedbrytningselasticitet’. Dessa skillnader i modellansats är inte triviala vad gäller skattning av marginalkostnader. De torde också ha kopplingar till andra typer av underhålls- och ombyggnadskostnader, även om ingen analys av detta har gjorts.

6. Ett viktigt antagande som ligger till grund för Newberys ’fundamentala teorem’ är att underhållsåtgärden är regelbaserad eller tillståndsreaktiv.

Normalt antas sålunda att när ojämnheten uppnår ett kritiskt värde kommer vägen att beläggas. Sådana principer följs också av väghållare, om än mer i teorin och planeringen än i verkligheten. Detta väcker frågan: om planeringsperspektivet antas vara relevant, och om, därutöver priser ska anpassas till marginalkostnaden så att de påverkar nuvarande och framtida vägutnyttjande, varför då inte använda de planeringsmodeller (PMS) som används av väghållarna, för att bestämma marginella infrastrukturkostnader. Idén är inte ny, men den har sällan använts. I översikten visas att denna ansats kan tillämpas på en specifik vägsträcka med hjälp av Världsbankens HDM 4 modell.

7. En implikation av att använda en PMS (såsom HDM 4) är följande. Dessa modeller bygger på antaganden rörande klimateffekterna och formen på sambanden mellan vägutnyttjande och förändringar i vägens tillstånd. Vad gäller beläggningar har visats att HDM 4 modellens antaganden är ungefär desamma som de antagande som finns i Newberys ansats (med implikationer som berör klimateffekten och ’nedbrytningselasticiteten’). Å andra sidan, genom att använda en PMS undanröjs behovet av att luta sig mot en partiell ansats vilket utmärker såväl den indirekta metoden som den ekonometriska metoden. Som visas med beräkningar som presenteras i denna översikt kan den mer heltäckande ansatsen vad gäller slitage och skador som normalt finns integrerade i en PMS, såsom HDM 4, mycket väl leda till slutatsen att trafikantexternaliteten inte borde ignoreras. Det finns ytterligare två aspekter som är värda att notera. Först, genom att använda PMS ansatsen kan marginalkostnaden relaterad till en viss fordonskategori beräknas direkt; det är inte nödvändigt att separat identifiera marginalkostnaden för var och en av de olika variablerna för vägutyttjande (givet att ESAL inte fångar alla effekter av vägutnyttjande). För det andra, ansatsen beaktar direkt alla typer av underhållsåtgärder, dvs. löpande, periodiskt och ombyggnad etc., som följer av vägutnyttjande. Andra fördelar är att ansatsen utnyttjar enhetskostnader för insatsfaktorer i vissa arbetsprocesser, vilka kan anpassas – genom kalibrering – till att återspegla förhållandena i ett speciellt land eller delar av ett land, inklusive den specifika utformningen av vägar vad gäller beläggningstyp och underbyggnad, etc. Dessutom, genom att använda en PMS kommer det att vara möjligt att bestämma marginalkostnader på olika aggregeringsnivåer, t.ex. för en specifik väg eller för ett genomsnitt av ett vägnät. Newberys ansats kan enbart användas för ett vägnät.

Ett antal skattningar av marginalkostnader, ibland uttryckt i termer av ESAL, ibland per km för olika fordonstyper, finns redovisade i denna översikt. Inga försök har gjorts att jämföra resultaten med tanke på att olika ansatser använts, formen och kvaliteten på de data som utnyttjats och att många empiriska resultat hänför sig till situationer som inte är relevanta för de sammanhang i vilka marginalkostnadsbaserad prissättning skall tillämpas. Denna översikt visar att det är möjligt att skatta marginalkostnader av vägutnyttjande, men kvaliteten kommer att variera. Innan sådana estimat används, måste dock kravspecifikationerna för hur de ska användas formuleras.

1 Introduction

The National Swedish Road and Transport Research Institute (VTI) has been commissioned by the Swedish Agency for Innovation Systems and the Swedish National Rail Administration to carry out a study of the marginal cost of transport. The study is expected to generate results of use for implementation of marginal cost based pricing of transport in terms of the principles laid down in the white paper on a common transport policy presented by the EU Commission in September 2001[2].

As concerns road traffic, a number of issues – related to the calculation of charges based on marginal costs – are being reviewed, including the nature of road deterioration and the so called rule of fourth power. In addition, VTI intends to carry out an international survey of approaches and methods used to the estimation of the marginal cost associated with road damage.

This report comprises the findings of this survey (the Survey), and covers not only road damage but also road wear and so called road damage externalities. These three marginal cost components are referred to as marginal infrastructure costs in this report.

As part of the work on the Survey, the following has been done:

1. A review of documented research (mainly published in English) since about 25 years, including reports and published articles. The emphasis has been placed on the methods used or that could be used to estimate marginal infrastructure costs. A review has also been made of other types of cost-based studies, to assess their relevance for generating data and information of use to the charging of road traffic.

2. An identification of – relevant – ongoing research with an emphasis on the EU, but also other OECD member states.

3. An analysis and evaluation of different approaches and methods for estimating marginal infrastructure costs. This has also included an evaluation of to what extent pavement management systems (PMS), in particular the World Bank sponsored HDM 4 model, may be used to calculate such costs.

4. Presentation, analysis and evaluation of empirical estimates of marginal infrastructure costs.

5. Identification and analysis of issues, and recommendations, with respect to forthcoming empirical work in Sweden to determine marginal infrastructure costs in road traffic.

A number of empirical estimates of marginal costs will be referred to in this report. As these have been estimated based on often very different assumptions, and also refer to different years, different countries and places, different currencies and different measures of road use, no attempt is being made to bring them together in one table. This also reflects that the emphasis of this report is on methods, and not on the results.

[2] Commission of the European Communities (2001).

2 Background

2.1 Why?

Roads are normally financed by way of appropriations. The exceptions to this are toll roads, as well as private roads. Another exception is roads operated by road associations. A limited number of countries also finance part or even the whole of the road sector by directly charging road users for use of the public road network. Whilst the roads are not private, the revenues generated by road user charges are not a part of general tax revenues, and are subject to expenditure rules that differ from those applying to revenues raised by way of taxes.

More or less all countries impose various forms of taxes on vehicles (in addition to taxes like VAT imposed on the sale of goods and services, in general). These taxes comprise specific taxes on fuels used by motor vehicles and licence fees (fixed fees paid on an annual or other time-related basis), which vary with type of vehicle, and characteristics such as gross vehicle mass (GVM), number of axles, size of engine, etc. A few countries, plus a number of states in the US additionally impose a weight-distance tax, normally on heavy goods vehicles only. A similar type of tax/charge, in use in the EU, is referred to as the Eurovignette. It has hitherto been applied only on heavy trucks, and on a time basis, and for the use of motorways and similar type of roads. However, Germany will convert its Eurovignette as from August 2003 to become distance-based. The new charge will be imposed on domestic and foreign large trucks and only for use of the motorways in Germany. Austria and Holland are also planning to convert their time-based Eurovignette to become distance-based.

Whilst the determination of the level of different taxes on road users ultimately is a prerogative of the parliament, increasing attention is being paid to this level for a number of reasons, including (i) efficiency in resource use, (ii) to ensure fairness between road users, and (iii) to limit the redistributional effects between road users and non-road users. From an efficiency point of view, increasing attention is being paid to the link between charges and the concept of marginal cost. This is not only because of a desire to promote efficient use of road capacity, but also to ensure that road users pay the marginal external effects of road use. This thinking is central to the EU transport policy, which promotes the concept of aligning taxes and road user charges to the short-run marginal cost as a key strategy to ensure efficiency. Emphasis on the short-run implies focusing on the direct resources consumed by an additional vehicle on a road, given that the capacity and the geometric features of the road are fixed.

This Survey deals with one particular aspect related to short-run marginal cost of road use, viz. the measurement of marginal infrastructure cost. Except where necessary, attention is not paid to congestion and the effects of emission and accidents and similar.

2.2 Roads, road use and road preservation

For the purposes of this Survey it may be useful to briefly review the basic terminology related to roads, including the effects of the use of roads by vehicles, and different actions related to the preservation of roads. The Survey will focus on paved roads, but without paying attention to structures and furnishings such as bridges, tunnels, drainage, barriers and road markings.

In general, road pavements are composed of a number of horizontal layers. These consist of bound materials on the top of the construction and unbound materials in the lower layers. Pavements are classified as rigid or flexible depending on the stiffness of the materials used. The primary function of a pavement is to spread the wheel loads, to prevent overstressing of the underlying soil and to provide a smooth ride surface.

In Sweden, the top layer of a flexible pavement, the wearing course, normally comprises about 40–50 mm of bitumen bound material. The second layer, the bitumen bound road base, varies between 40–190 mm depending on traffic intensity. The thickness of the third layer, the unbound road base, varies between 80–150 mm depending on the required bearing capacity and the quality of the gravel material. The fourth and last layer, referred to as the sub-base, comprises uncrushed gravel of between 350–420 mm. The sub-base rests on the sub-grade.

In Sweden rigid pavements are recommended only for high volume roads. They normally also comprise four layers as for a flexible pavement, but the wearing course is made of concrete. The road base may be of bitumen or cement bound material.

The use of a road results in various forms of distresses. Both flexible and non-flexible pavements can be subjected to distress from above, but also by motion of the underlying soil, which is affected by stresses on the pavement itself and by moisture and temperature conditions underneath.

Pavement defects can be classified as surface distress and deformation distress. The latter affects the structural condition or the pavement's bearing capacity. Reduced bearing capacity results in faster deformation. Deformation may be observed by way of spot settlements, increase in roughness and development of rutting.

Surface distress manifests itself in cracking, ravelling, potholes, edge breaks, effects on the surface texture (which can be measured by way of the texture depth and the skid resistance), rutting and increased roughness, i.e. longitudinal unevenness.

Deformation reflects damage to the pavement and its base on account of the load on the road[3], whilst surface distress can be viewed to represent the wear of a road on account of the passing of vehicles. Wear affects primarily the surface of the pavement, and also results in reduced friction. Impaired friction affects road users in that it increases the risk of accidents. If cars have studs during the winter, they may result in rutting. However, rutting is normally the result of damage by heavy vehicles. Rigid pavements tend to crack rather than deform when subjected to stress.

However, roads are not only affected by traffic. Even if a road is not used at all, drainage functions (i.e. water is not allowed to seep into the road, but flows off the pavement and away from the roadway) and weeds along the road are kept in place, it will deteriorate[4]. In addition, the wear and damaging impact of a vehicle is magnified by the climate (the weathering effect). In particular, the freeze-thaw cycle is an important reason for pavement deterioration in countries with temperate climate.

---

[3] Wear is here used in a different way than what is common in Sweden, where this term primarily refers to the effects of studs.

[4] This is more true for flexible than for rigid pavements.

All types of vehicles give rise to wear, whilst traffic-related damage is principally caused by the pressure of the wheel against the pavement, i.e. by loads, and hence by heavy vehicles. Wear – if not controlled by way of maintenance – can eventually accelerate the deformation process. Damage is, as mentioned, *inter alia,* measured by way of roughness, but certain types of surface distresses also add to increased roughness. It is common for roughness to be measured by the International Roughness Index (IRI).

To measure the damage impact of different types of vehicles, the concept of equivalent standard axle loads (ESALs) is made use of[5]. One ESAL corresponds to the damage caused by an axle with two wheel pairs with a load of 100 kN equally distributed on the wheels. The relationship between the damage of one ESAL and that of another load is given by the following formula:

$$\text{ESALs} = (x/100\text{kN})^{\omega} \qquad (2.1)$$

where x is the axle with unknown ESALs, the load of which may be measured in kN. In Europe the weight of a standard axle is assumed to be 10 tons, whilst a standard axle in United States is 18 000 pounds (about 8.2 tons)[6].

It is frequently assumed that ω is equal to 4, but this is subject to debate as discussed further below. Given the number of axles, wheels and GVM of a particular vehicle, it is possible to establish its Vehicle equivalence factor (VEF), which indicates the total number of ESALs associated with that vehicle when it is fully laden, and hence its total damaging effect. The VEF of a laden heavy truck is normally about 10 000 higher than that of an ordinary car.

Roads are often designed to withstand a certain number of ESALs, i.e. the design life is measured in terms of ESALs. The higher the traffic and the more heavy traffic on a road, the thicker the road, not least the pavement, should be from an economical point of view. The design life of a road in terms of ESALs can be seen as a measure of its structural condition. However, the structural condition is often measured in different ways, e.g. by reference to the so called structural number of the pavement.

Roads are often seen as ever lasting capital objects. However, to ensure their longevity a number of actions have to be taken in order to preserve them. These include maintenance which affects the parts of the road that are visible. When maintenance is not adequate, reconstruction will be required. The following terminology is being made use of here:

1. **Routine** maintenance comprises two types of actions. The first type includes vegetation control, maintenance of the ’furniture’ of the road and clearing of drains and culverts to ensure the run-off of water, and a functioning drainage system at all times. The second comprises the sealing of cracks, the patching of potholes and ravelled areas, and the repairing of edge damages. Routine maintenance is carried out on a regular basis every year, and have effects which normally do not last for more than a year. However, whilst the quantity of work related to the first kind of routine maintenance may be seen to be independent of the amount of traffic, the

[5] Most of the design procedures in use today are also based on the concept of an ESAL.
[6] If it is assumed that ω is equal to 4, then a European (EU) ESAL is about 2.21 a US ESAL.

resources spent on repairs of cracks, potholes, etc., may be seen to be a function of the volume of traffic, and the amount of heavy traffic. A policy of road authorities with respect to these works is normally to ensure that the amount of surface and edge damage is kept within a certain level.

2. **Periodic** maintenance includes two types of actions. Firstly actions which result in improved texture and smoothness, but does not enhance the structural condition of the pavement (e.g. fog seal and single surface dressing; these actions are also referred to as 'resealing'). Secondly actions in order to enhance the structural condition of the pavement, i.e. overlays, which comprise an addition of a thick layer, and pavement reconstruction, which is a complete new overlay. Resealing removes the small excess of roughness due to surface defects, and forestalls more rapid pavement deterioration, but has essentially a small effect on roughness. Overlays result in a major reduction in roughness, and are substantially more costly and less frequent than reseals. Decisions on periodic maintenance actions are normally taken in response to the condition of the road, as discussed further below.

3. **Reconstruction**, which is done at the 'end of the life' of the road, when parts of the base have to be reconstructed, in addition to the pavement. Decisions on reconstruction are also normally condition responsive.

Development in pavement technology implies that a primary road (normally with an asphalt pavement) being built today will only be subjected to a form of overlay after some 10 years, and then again after perhaps 10 years, and so on. Routine maintenance more and more only involves drainage and furniture only, whilst pavement maintenance of a routine nature is not expected to be required. Resealings as a maintenance action are also becoming a rare phenomenon, and modern roads are not expected to need reconstruction even in a long term perspective. It is also expected that pavements and overlays gradually will come to last longer and longer on account of better quality of the bitumen. However, it should also be mentioned, that whilst overlays are becoming ever more dominant as maintenance actions, there are different types of overlays, depending on the nature of the damage and also the age of the pavement. For example, damage on account of studs may be handled by one type of overlay, whilst damage caused by heavy vehicles precipitate a different type of overlay (see further below in Section 3.2).

In addition to maintenance, roads also require inputs in the form of **operations**, comprising winter maintenance and electricity to power lights and ventilation systems in tunnels, etc. These inputs are (mainly) independent of traffic.

2.3 The effects of road use

The presentation in the previous section has indicated that distresses are caused by traffic but in different ways. The implication hereof is that road use will have to be measured in different ways, and the following definitions are therefore provided of road use during a given period of time (e.g. a day or a year):

1. v-km, i.e. the traffic measured as the AADT (if the length of the period is one average day) multiplied by the length of the studied road,
2. axle-km, i.e. the number of axles multiplied by length,
3. GVM-km, i.e. the total gross-vehicle mass (the weight of the vehicle and its load) multiplied by length,
4. ESAL-km, i.e. the total number of ESALs multiplied by length,
5. PCU-km, i.e. to total per car equivalent units multiplied by length.

PCU is a measure of the space requirement (e.g. a normal truck is equal to three cars and therefore has a PCU value of 3), and is often used to allocate congestion and capacity costs in cost allocation studies (see Section 5). Sometimes a more refined indicator of space requirement (PCE), which also takes into account the difference in normal speed between vehicles, is made use of. If these indicators of traffic or road use are referred to without ’km’, then reference is made to a road length of one km.

A marginal vehicle on a road can be measured in terms of each one of these indicators. Of course the four last measures are primarily of interest to catch the effects of heavy traffic, in particular goods vehicles.

The short-run marginal infrastructure cost associated with an additional vehicle on a road comprises three components. These are:

1. The increase in the cost inflicted on other vehicles as a consequence of the additional vehicle. This cost, referred to as the **road damage externality**, reflects that the vehicle will cause some wear and damage to the road, and that this deterioration will result in increased costs to subsequent vehicles in the form of either (i) increased cost for operating the vehicle, (ii) lower speeds, resulting in higher time costs, and/or (iii) less comfort when driving the vehicle[7].

2. The fact that additional wear of the road will lead to the road authority taking action to remedy the road **wear** by way of **routine** maintenance actions (**the marginal cost of road wear**) at an earlier data than would have been the case without the additional vehicle. It is assumed that the road authority monitors road wear and takes action when the effects of road wear, e.g. in the form of reduced friction, results in a condition worse than a certain critical level. For there to be a marginal cost associated with road wear, it has to be assumed that maintenance actions with respect to road wear are condition responsive.

[7] These costs are the costs borne by road users, and should in principle also include the increase in risk of road accidents, although this cost component is not reflected in models used today, neither for planning nor prediction purposes.

3. The fact that additional damage of the road will lead to the road authority taking action to remedy the **damage** by way of periodic maintenance actions or reconstruction at an earlier date than would have been the case without the additional vehicle. It is assumed that the road authority monitors road deformation and takes action when the effects of damage, e.g. in the form of increased IRI value, reaches a certain critical level. It is thus assumed that maintenance actions with respect to road damage is condition responsive (**the marginal cost of road damage**).

It is clear that the three costs associated with a marginal user will have to be a discounted (or a capital) value, as a marginal vehicle will give rise to costs which take place not just at the same time of the use of the road by the marginal vehicles, but during all future periods. This condition identifies an issue related to the measurement of marginal cost of road use, viz. the level of the discount rate to be applied.

2.4 An overview of different approaches to the measurement of marginal infrastructure costs

Four different approaches to the measurement of the marginal cost of road use can be identified. The first will be called the **direct** method. Whilst it appears to be the most obvious and, in a sense, the simplest method, it has hardly been used at all for reasons unknown to us. The direct method is characterised by that it yields estimates in the form of a discounted value and does not necessarily make a distinction between the three components. The second method will be referred to as the **indirect method**. It is based on the 'fundamental theorem' formulated by Newbery[8]. The characteristic of this approach is that it is based on the assumption that the road damage externality cost may be ignored. It focuses on the consequences of road damage and periodic maintenance in the form of overlays. Limited work has been made to extend this approach to apply to road wear and condition responsive routine maintenance, and reconstruction. There exists, however, one frequently used approach to the measurement of the marginal cost of road use which could be seen as an extension of the Newbery approach, also considering wear and condition responsive routine road maintenance. This is the RUC30 model developed by the World Bank[9].

The majority of road costing exercises to be found have no clear foundation in the marginal cost approach although they contain results which could be seen to be, and often are, of relevance to the estimation of marginal costs. These cost studies, frequently referred to as **cost allocation studies**, essentially have a different perspective in that they focus on equity, and (the aggregate of) variable and fixed costs of road works (i.e. costs of road works which are viewed as being a function of traffic, measured in terms of one or several of the above measures of road use, and not being a function of traffic) as opposed to marginal and non-marginal costs associated with an additional vehicle. In this Survey these cost allocation studies will also be referred to as belonging to the club approach as estimates from them can essentially be seen as trying to answer the following question: Assume that all users of a road belong to a club, and that they have to

[8] See Newbery (1988b); (1988c) and (1989).

[9] See Heggie and Vickers (1998) and Archando-Callao (2000).

agree on a system for how to recover the costs of road works by way of user charges. What would then be the characteristics of such charges?

The three previous approaches are based on estimations essentially making use of various types of unit costs (e.g. the cost of an overlay per km, etc.). The fourth approach also relies on unit cost of sorts, but marginal costs are determined first after having estimated a cost function by way of **econometric** techniques. This approach builds on conventional microeconomic production theory, but there are few examples of it.

3 The Direct Approach

3.1 Introduction

The idea of the direct approach is to make use of a pavement management system (PMS) model for estimating the marginal cost of road use. The most common PMS is the HDM 4 model, used in developing countries and other countries in which international financing institutions operate[10]. HDM 4 may be used in order to analyse the economic effects both of various maintenance policies and of investments in the road network. The Swedish road authority makes use of a similar PMS model, but primarily for determining the appropriate road maintenance strategy and activities.

The output of an analysis using a PMS model is typically an estimate of a net present value (NPV) of a certain action. Normally, two alternatives are compared, and the difference in the NPV between these two alternatives calculated. The cost of a marginal user may be determined, in principle, as the difference between the NPV of two alternatives, one having an increase in the AADT by one unit in comparison with the other. To obtain the marginal cost of one vehicle, corrections will then have to be made to account for the fact that the analysis is based on an increase in one unit every day, and during all days in the future, and not just one unit on one particular occasion. It is in this way, in principle, possible to determine the marginal infrastructure cost for many different types of vehicles, from ordinary cars to heavy trucks with, say 6 axles and 22 wheels. See the example below.

A marginal cost calculated by way of a PMS will normally directly reflect the three marginal cost components of road use. It may also contain further elements, assuming that they are relevant, viz. the marginal cost of congestion and of various external effects, provided the analysis is structured to include these components. Estimation can be made at different aggregation levels of the road network in order to match the requirements of the marginal cost-estimate. It is thus possible to estimate the marginal cost for a specific unit of road or to reflect the average marginal cost-value for a road network. It is also possible to estimate marginal costs for a given road (or network) at different points in time to reflect how the marginal cost may vary over time for a specific road.

Marginal costs estimated in this way suffer from the many weaknesses that characterise estimates of marginal cost obtained with all four methods considered here, including cost allocation exercises. These problems are e.g. the (i) the assumptions made with respect to non-traffic impact on road deterioration, (ii) the size of the coefficient ω in formula (2.1), typically assumed to be 4, and (iii) the impact of new technologies, e.g. the effects of introducing long-lasting pavements. Some of these generic problems will be considered briefly further in Section 7.

But the approach has some advantages in comparison with other approaches. Firstly it is based on a model of either an individual road or a road network that may be used to determine the appropriate level of maintenance and investment in the road or the road network. The approach thus provides a direct link between investment, maintenance and pricing. Models like HDM 4 are, in addition, increasingly being used for planning and programming purposes by road authorities. Secondly, a PMS will normally include comprehensive and simul-

[10] See HDM 4 Manual (2001).

taneous modelling of all the effects related to increased road use, including effects on other vehicles, road wear and damage. And thirdly, these models are often based on up to date empirical relationships between traffic and the effects on roads and other vehicles[11].

One drawback of the approach is that the results of the calculations are not transparent. To examine the estimated result it will be necessary to analyse a series of calculations involving several relationships, and these calculations are not presented as a report of a run of the model (e.g. HDM 4). Better transparency would require modifications to the software. A second shortcoming is that models of this nature cannot be applied in a new environment without comprehensive calibration, to ascertain that local conditions are reflected in the empirical relationships.

3.2 An example to illustrate

The table below contains the results obtained from a run with HDM 4 for a specific road. The road is 9 m wide and carries an average daily traffic of 6,000 vehicles, including 120 buses, 360 large trucks, 360 small trucks, 4,860 cars and 300 pick-up vehicles. The calculations have been made for a 50 year period, a real rate of interest of 4%, and the road is assumed to be newly constructed at the beginning of the period. The maintenance strategy is that that the road is subjected to an overlay if IRI exceeds 5.0, a partial overlay if rutting exceeds 22 mm or a partial resurfacing if more than 10% of the road area is affected by wide structural cracking. These are maintenance strategies that may be considered for roads in southern Sweden.

The table sets out the calculated marginal costs for each type of vehicle, in SEK per km, and the marginal cost is also divided into two components, i.e. on account of wear and damage, and the road damage externality. As may be observed, the latter component is much more important than the former. Calculations of the marginal costs with respect to these two components for an average road, i.e. a road which is in the midst of the maintenance cycle over time, indicate that these costs will differ from those of – but will be of a similar magnitude as for – a new road[12]. The size of the road damage externality also implies that MC-based pricing will likely more than recover the costs associated with the wear and damage to the road during a 50 year period. It is to be mentioned that the road damage externality does not reflect congestion costs in the calculations presented here.

[11] The empirical relations included in the HDM 4 model (and similar models) will not be covered here. See, however, HDM 4 Manual (2001) as well as Watanatada et al. (1989).

[12] As will be discussed later in this report (see Section 4.1), under certain conditions the road damage externality can be assumed to equal zero. Indeed, most calculations of marginal costs of road use is based explicitly or implicitly on the assumption that the road damage externality may be ignored. The calculations presented here suggest that this may not be an appropriate assumption, but the exact reason is unclear. It is, however, noted that the empirical relationships embodied in the HDM 4 model are much more comprehensive than those of other models used to calculate marginal costs.

Table 1 *Marginal cost calculations with HDM 4; SEK per km.*

Vehicle	Wear and damage	Road damage externality	Total
Bus	0.001635	0.516679	0.518314
Truck	0.015533	0.741432	0.756965
Light truck	0.000204	0.134756	0.134961
Car	0.000177	0.026454	0.026630
Pick-up	0.000000	0.064340	0.064340

Source: Own calculations

3.3 Simplified direct approach

In the above example, a condition responsive maintenance policy with regard to wear and deformation has been assumed and applied, implying that a marginal vehicle affects the timing of certain routine and periodic maintenance actions. This may not always apply. If maintenance is not condition responsive, the marginal cost will only comprise the road damage externality. The cost of this externality will then have to be calculated in order to obtain an estimate of marginal cost, and the study by Gronau (1991) provides an example of how this may be done. His approach is based on making use of the type of empirical relationships included in HDM III (the predecessor of HDM 4) to determine in a first step the impact of one additional vehicle on roughness over time, and in a second the consequence of this increased level on roughness on the cost of vehicle operations (but excluding the time costs of travel) during following periods. Calculations of this nature can also be made with HDM 4, by running the model subject to maintenance actions being non-traffic responsive, but to be effected e.g. after fixed periods of time.

In Gronau's study, making use of data for Ghana from the 1980s, an estimate of 0.9 US cents per US ESAL-km was obtained. This corresponded to 1.53 to 3.87 US cents per v-km for heavy vehicles. Similar estimates were obtained by using other approaches to the estimation of marginal costs, e.g. by making use of the indirect method considered below[13].

[13] These results are also reported on in Gronau (1994), which additionally includes estimates for Zimbabwe. It is, however, unclear which method of estimation that was used for the latter country.

4 The Indirect Approach

4.1 The theory

The Indirect Approach is based on the same approach to the identification of costs associated with a marginal user as for the direct approach. There are some differences, however.

- It defines the marginal vehicle as an increase in vehicles by one vehicle not just during one period, but also for all subsequent periods. Assuming a period to correspond to a year, then the marginal vehicle refers to an increase by one vehicle during all subsequent years as well[14].
- It focuses on road damage and road overlays. Overlays make up a substantial part of the total costs of maintenance and operations of roads, and as mentioned in section 2.3 this dominance is being reinforced by developments in pavement technology and maintenance practices.
- It ignores the wear on roads.

The 'fundamental theorem of road damage' formulated by Newbery (1988b) and (1989) states that the marginal cost is proportional to the (average) cost of (overlay) maintenance per ESAL-km, or

$$MC = (\Theta x C)/(T x Q) \qquad (4.1)$$

where C is the cost per km of overlay, T is the number of years between two overlays, Q is the annual traffic in ESALs and Θ is the share of road deterioration explained by traffic. If this share is equal to one, then the marginal cost is simply the cost per km of overlay, divided by the accumulated number of ESALs during the entire life of that overlay.

As shown by Newbery, this formula applies subject to the following conditions:

- The road network has a uniform age distribution (or each type of road has a uniform age distribution).
- Road overlay maintenance actions are triggered by road conditions (Newbery considers a maximum level of roughness as the indicator of when to act).
- There is no traffic growth.
- Damage is only caused by traffic (i.e. Θ is equal to one, see below).

The importance of these conditions is that they imply that the cost associated with the road damage externality becomes equal to zero. This is essentially a consequence of uniformity and the fact that the marginal vehicle causes an overlay to take place earlier than envisaged originally, which means that whilst the cost of the externality for traffic immediately after the marginal vehicle will increase (in comparison with the case with no marginal vehicle), it will later decrease since of the timings of overlays are being moved forward. It may be shown that formula (4.1) is a good approximation also in the case of positive traffic growth over time and the road is damaged not only by traffic but also by

---

[14] The difference is not significant, as the calculations made in the example in Section 3.2 in effect is based on the same type of assumption.

the weather. The weathering effect results in the coefficient Θ taking on a value between 0 and 1.

The relationship (4.1) is derived from the assumption that, given a condition responsive maintenance policy, and an assumed empirical relationship between Q (measured in ESALs) and the level on the measure of distress that triggers an overlay, the following applies to the timing between overlays (T)[15]:

$$T = \bar{R}S^{\sigma}/(kQ + zW) \qquad (4.2)$$

where S is the structural condition of the road, σ, k and z are empirical constants, W is the weather effect (e.g. the annual precipitation), and

$$\bar{R} = R_o - R^*$$

R^* is the critical value on the pavement condition rating resulting in an overlay and R_o is some initial rating. Now, given this formula, it may be shown that

$$\Theta = kQ/(kQ + zW)$$

So if the weather effect can be ignored, then Θ becomes equal to one, which means that the marginal cost associated with an overlay is the cost per km of overlay divided by the accumulated number ESALs. Normally when applying this approach to estimating the marginal cost of an overlay, roughness as measured by IRI is used as an indicator of pavement distress.

The are a number of variants of this approach to the estimation of marginal costs of road use. The approach used by Lindberg (2002) differs from that of Newbery in that it is based on a different relationship for determining the timing of an intervention. Firstly Lindberg makes use of a cracking index to determine intervention, and secondly the property of the relationship between time and the critical level of distress, triggering an intervention, is of a different nature. Additionally, the empirical relationship for the cracking index uses only traffic as an explanatory variable (in addition to structural condition)[16]. The change in T for a small increase in Q for (4.2) is as follows:

$$dT/dQ = - kT(kQ + zW)$$

Deriving the elasticity of T with respect to Q (the deterioration elasticity in Lindberg's terminology), and assuming that zW can be neglected (as done by Lindberg), the following is obtained

$$\varepsilon = QdT/dQT = 1$$

i.e. the elasticity is unitary. The same does not apply for the relationship between time duration and the pavement indicator used by Lindberg. He consequently writes his relationship for the marginal cost as

[15] Cf. Vitaliano and Held (1989).
[16] This relationship has been derived from data, suggesting no weathering effect.

$$MC = \varepsilon C/(TxQ), \quad (4.3)$$

where ε is the deterioration elasticity. In Lindberg's empirical work this elasticity is of the order 0 to 1, so that the marginal cost is lower than the average cost.

Another formula is offered by Vitaliano and Held (1989). Their short-run MC is written as

$$MC = \alpha \Theta CT/Q \quad (4.4)$$

which is the same as (4.1) if $\alpha=1/T^2$. The α factor in the formula of Vitaliano and Held is an annuity factor, and it is shown in their study that its empirical value is almost the same as $1/T^2$. It is believed that the difference in the formulas is explained by the fact that Newbery is considering an average road (at an average time in-between overlays), whilst Vitaliano and Held consider a road which is about to be overlaid. Apparently, the differences are not material. The Vitaliano and Held model is the same as the one used by Transportation Research Board (1996) in the study Paying our Way[17], as well as Small et al. (1989), although in the latter study Q is replaced with TQ/T, where TQ is the cumulative number of ESALs until the next overlay. The Small et al. model also differs from Newbery's in that the weathering effect is shown to raise the value of the marginal cost to become higher than the average cost. Although both Newbery and Small et al. make use of the same source as to the empirically established relationship between roughness progression and traffic measured in ESALs[18], they arrive at different results as to the relationship between the marginal cost for an overlay and the average cost, apparently on account of different mathematical formulations of this relationship.

4.2 Extension of this approach

The fundamental theorem applies to overlays or periodic maintenance activities. Although no formal work has apparently been done, routine maintenance active-ties are at times handled in a similar manner as periodic maintenance and the marginal cost associated therewith[19]. This extension of the Newbery approach seems to be based on the assumption that the wear and damage of roads as witnessed by cracks, potholes, edge-wear, etc., give rise to condition responsive routine maintenance activities, which may be analysed similarly to the overlay costs, in that the road damage externality may be ignored. The implied assumption is that this component of the marginal cost can be determined as the routine maintenance cost per km divided by some measure of road use.

In his empirical work on road costs in Britain, Newbery (1988a) and (1990) has e.g. allocated costs for reseals and repairs of potholes on the basis of v-km and GVM-km, respectively. And in his seminal work on marginal cost-based user charges, i.e. the study for Tunisia, Newbery allocates resealing costs due to ravelling based on axle-km and resealing costs related to cracking based on

[17] A similar approach had been used in two previous studies by the Transportation Research Board in 1990.

[18] The source is the empirical work done as part of the development of the HDM model as reported on in Watanatada et al. (1989).

[19] See, however, Newbery (1985).

squared ESAL-km[20]. Although not explicitly stated in these reports, it is indicated that the same approach as used for overlays may be applied to routine maintenance works, given that the wear and damage are explained by different measures of road use, and that the weathering effect may be ignored[21].

4.3 The RUC30 model

The most widely used model to determine marginal cost-based road user charges is the RUC30 model developed by the World Bank. It is a simplified model, based on the theoretical and empirical work which has gone into the development of HDM 4 and the Newbery approach for how to calculate marginal infrastructure costs. The RUC30 model, including software and manuals, can be downloaded from the World Bank homepage, as can several reports illustrating its use[22].

The RUC model is based on a separation between administration, routine maintenance and periodic maintenance costs, on an annual basis. Routine (or annual) maintenance and periodic maintenance costs are divided into fixed and variable costs. The fixed cost element is supposed to reflect the weathering effect, whilst the variable costs reflect the impact of traffic. A differentiation may be made between different classes of roads, reflecting traffic level, condition and geometric standards. The model can handle a number of different types of vehicles.

To compute marginal costs, it is necessary to provide cost estimates on an annual basis for routine and periodic maintenance activities, both the element explained by weathering and the one explained by traffic. But default values can also be provided by RUC30. These default values can also be determined by a separate model. This supplementary model relies on inputs in the form of data on vehicles, vehicle operating costs, climatic environment, vehicle loadings (ESALs)[23], the traffic on the road (or type of road) studied, the pavement structural condition, and unit costs for various types of maintenance activities (e.g. overlay per sq. m. etc.). The supplementary model then determines optimal routine and maintenance costs, including their division into traffic and non-traffic dependent costs. The optimal costs correspond to an assumed condition-responsive maintenance policy; they have been determined through a large number of model runs by way of the HDM 4 model for assumed standard cases.

In determining marginal costs, the RUC model allocates traffic dependent periodic maintenance costs based on ESALs. All other variable costs are allocated based on vehicle-km. A way of describing the RUC30 model is thus that it is a simplified approach to determining marginal costs by way of a planning model, in this case the HDM 4 model[24]. The simplifications embodied in RUC30 are essentially based on the Newbery approach, which allows traffic dependent

[20] See Newbery et al. (1988).

[21] A neglected aspect is the interdependence between maintenance actions. For example, an overlay obviates the need for routine maintenance. Presumably, thin overlays also obviates the need for some other routine maintenance activities.

[22] www.worldbank.org/transport/ ; see also footnote 6.

[23] Input data are in the form of equivalent axle load factors for different vehicles, which means that the use of RUC30 is not tied down to the 4th power rule or assumptions about the weight of a standard axle.

[24] Indeed, RUC30 can also be used to estimate other marginal cost elements, reflecting accidents, emissions, and congestion albeit in a stylised fashion.

maintenance costs to be allocated in terms of various measures of road use. The model makes use of only two measures, however, viz. ESAL-km and v-km.

4.4 Empirical results

This section mainly includes results from studies focusing on short-run marginal cost estimation. Additional empirical results are given below where the marginal infrastructure costs have been estimated as part of total cost allocation studies.

4.4.1 Swedish studies

Following the government's decision in 1979 to focus on marginal costs as a basis for pricing of infrastructure in Sweden, road cost studies have essentially been limited to the estimation of marginal costs. Studies in the 1980s and early 1990s were based on an equivalence between variable maintenance costs and marginal costs; these studies are not commented upon here[25]. More recent studies are more elaborate.

A study by the Swedish National Road Administration (Vägverket (2000)) focuses initially on the allocation of total maintenance and operations costs on variable and non-variable cost elements. Total annual maintenance and operations costs have been estimated on the basis of assumptions about what is required to maintain the road network at a steady state, an assumption which must be seen to reflect a condition responsive maintenance policy. Variable costs are seen as road wear and damage costs caused by vehicles. The variable costs are subsequently split up into two categories, viz. those which are explained by heavy vehicles and those explained by ordinary passenger cars. The allocation of costs is made for different road classes (defined in terms of AADT), and the split first between variable and non-variable, and subsequently between costs depending on passenger cars and costs explained by heavy vehicles is done with reference to experience. These allocations are hence not based on an a more formal analysis of the nature of distresses resulting in a conclusion that different maintenance activities are explained by different measures of road use. The study mentions, however, that the assumed distribution of costs has been gauged against the results of runs with the PMS used by the Administration (a model similar to HDM 4) to ensure that they are reasonable. Variable costs for heavy vehicles are allocated on the basis of ESAL-km and variable costs for cars are allocated per v-km, vehicle equivalence factors are then used to convert costs per ESAL-km to a marginal cost for different types of heavy vehicles.

The findings of this study was that the marginal cost for an ordinary car was about 1 öre/km, whilst that of a heavy truck was about 15 öre/km.

The study by Lindberg (2002) is explicitly based on the Newbery approach to the estimation of marginal costs. It provides hence a justification for using average costs per ESALs. The study complicates the Newbery approach by assuming that the deterioration elasticity is not unitary (see formula (4.3) above). The empirical work with this approach focuses on overlays (i.e. periodic maintenance), and is based on the assumption that the weathering effect can be ignored[26]. Lindberg illustrates how his approach can be used to determine

[25] These studies include Ds K 1985:2 & 3, Ds K 1987:11, Ds 1992:44 and SOU 1996:165, which were all sponsored by the government (directly or indirectly); see SIKA (2000).
[26] See footnote 14.

marginal costs, as well as how these costs vary with respect to the original structural condition of the road, its width and traffic. The National Road Administration in a subsequent report (Vägverket (2001)), in principle accepts the approach of Lindberg to measure marginal costs of overlays. The report, however, also points out that the empirical data used by Lindberg are not representative for the road network as a whole in Sweden, although his findings are, on average, not all that different from the estimates produced in the Vägverket (2000) report. The Administration is also puzzled by the fact that the marginal cost of an ESAL is higher for roads with heavy traffic than for roads with little traffic; these latter roads normally have weaker pavements[27].

The average marginal costs estimated by Lindberg amounted to between € 0.3 cents to € 1.9 cents per km for different types of goods vehicles. These results are also presented in the 2001 report from the Swedish National Road Administration[28].

4.4.2 Estimates by Newbery

In his seminal work, with data for Tunisia, Newbery[29] found a marginal cost with respect to overlays in the amount of 0.80 US cents per US ESAL, the marginal cost of resealing on account of ravelling was estimated at 0.003 US cents per axle-km and the marginal cost of resealing on account of cracking was estimated at 0.052 US cents per squared US ESAL-km. These data apply to the 1980s

Applying a similar methodology on 1986 UK data, Newbery (1988a) estimated the marginal cost of an ESAL-km at 1.4 pence per km, the marginal cost of a GVM-km at 0.08 pence per/km and the marginal cost of a v-km at 0.065 pence per km. Using 1989/90 data for the UK, Newbery obtained a revised estimate of the marginal cost for overlays in the amount of 3.5 pence per ESAL-km; Newbery (1990)[30].

4.4.3 US Estimates

Small et al. (1989) found the cost per ESAL-km to vary between 0.9 and 78 US cents per km at the then ruling principles of investments. Assuming optimum investment principles, the range was estimated at 0.3 to 63 US cents per ESAL-km. Vitaliano and Held (1989) using a similar approach also found similar values, they estimated the marginal cost of overlays to vary between 0.7 and 17.5 US cents per ESAL-km depending on the functional class of the road. This corresponded to between 1.9 and 49 US cents per km for a heavy articulated truck with 18 wheels. Both these studies refer to the situation towards the end of the 1980s in the US. Whilst Small et al. use data for the entire US, Vitaliano and Held rely on data from New York State. The most recent study on marginal costs sponsored by the Transportation Research Board[31], does not provide any explicit information on the marginal cost per km with respect to road use variables or for different vehicles. This information is embedded in case studies for different types of shipments in the US.

---

[27] The explanation of this is unclear, and the data in the Lindberg and National Road Administration reports are confusing.
[28] See also SIKA (2002).
[29] See Newbery et al. (1988).
[30] It is unclear if ESALs have been measured in US or European units.
[31] See Transportation Research Board (1996).

4.5 Issues related to the measurement of marginal cost with the Indirect Approach

Whilst the above studies have a common theoretical base, there are also substantial differences between how estimates of marginal costs are obtained. The following should be noted:

Firstly, two alternative approaches for estimating the cost of maintenance activities, e.g. overlays, are used. One is based on actual recorded annual costs, the other relies on estimated costs based on unit prices. RUC30 is based on the latter, which was also used by Vitaliano and Held (and Lindberg?). Newbery used the former approach for the UK.

Secondly, the assumption made about the share of costs explained by weathering varies. Newbery assumes 60% for the UK, whilst Vitaliano and Held assume 50% for New York State. Transportation Research Board (1996) makes more sophisticated assumptions resulting in this share actually varying with traffic, and increasing towards one for large ESAL values. Small et al. (1989) argue that weathering effects raise marginal costs above average cost, whilst Lindberg ignores this aspect.

Thirdly, the assumptions made about the functional relationship for the time between intervals and measure of distress (e.g. roughness) vary As mentioned, most analysts assume unitary elasticity, whilst Lindberg uses an elasticity of between (-0.1 and -0.8) resulting in much lower estimates of marginal costs, cet. par.

Fourthly, most of the reviewed studies only consider the marginal cost with respect to an overlay costs (e.g. Lindberg), but sometimes other – routine – maintenance costs are also included, for example in the Newbery studies. These later costs are also catered for in the RUC30 model. A cost element generally not referred to in the empirical studies are costs for reconstructions.

Fifthly, the marginal costs are estimated at different levels of aggregation. For example Lindberg provides estimates for different types of roads, including with respect to traffic, geometry and structural condition, whilst Newbery provides aggregate estimates for entire road networks.

Sixthly, care has to be taken of how an ESAL is measured, i.e. whether with respect to a 8.2 or a 10 tonne axle. From the point of view of the marginal cost for a particular vehicle, the unit of measurement does not matter. But as noted, results are often reported on in terms of ESALs only, and for a number of the reported results it is unclear as to whether the US or the European definition of a standard axle has been used.

Seventhly, mention should be made of that all the afore-mentioned studies are based on assumptions of rule-based behaviour on part of the road authority. The precise nature of the assumptions made vary; most refer to IRI values, whilst Lindberg considers cracking. This issue has, however, wider significance as the relevance of the assumption of optimal behaviour by road authorities may be questioned. Road authorities are generally funded by appropriations and subject to weak accountability structures. It is not obvious that officials behave 'optimally' under such conditions.

And finally, note should be taken of the results presented in Section 3.2, which may suggest that the assumptions which underlie the direct approach, giving rise to a road damage externality cost that may be ignored, may not be appropriate.

5 Club and Equity Approaches

5.1 Introduction

Most existing cost allocation exercises are based on what may be called the equity approach. It includes the club approach. These approaches are not necessarily incompatible with the marginal cost approach; there are overlapping elements as we will see. The crucial distinction is, however, that cost allocation and marginal cost estimation start from different premises. The marginal cost-approach is concerned with the marginal user, whilst the club and equity approaches are concerned with the allocation of the total cost of roads to road users, and the related problem of how to recover those costs.

The equity approach is based on the principle of fairness. What is meant by fair is often not clear, but one principle that is normally applied is that road users should bear the costs of the road network, related to its provision and preservation. Another principle is a vehicle owner should bear the costs that his particular vehicle gives rise to. That cost is normally – but not always – seen as consisting of two components. The first is related to variable maintenance costs, i.e. costs for maintenance which are seen as dependent on traffic. The second element is fixed costs related to construction. Some of these fixed costs may be seen as being occasioned by specific types of vehicles. For example, the durability of a road must be made much stronger in order to withstand heavy traffic. If the road were built to meet only the needs of ordinary cars it would be much cheaper to build. Hence, many cost allocation studies are based on the premise that heavy vehicles should bear the additional – or incremental – costs required to allow them to make use of the road[32]. Road network costs which cannot be attributed or allocated in this way, including the fixed maintenance costs and other costs that cannot be allocated to specific vehicle classes in terms of this approach, are allocated according to some 'neutral criterion', e.g. per v-km driven. Similar criteria may also be used within vehicle classes to allocate costs between individual vehicles, e.g. costs which are considered to be occasioned by a specific type of heavy vehicle.

The basis for the equity approach is thus, that each type of vehicle should bear ’its’ cost incurred as a consequence of both an investment decision or a decision to make use of a road. Remaining costs should be allocated based on the principle that each vehicle or v-km should be seen as being of equal weight.

[32] This appears to be the conventional approach. But there are other approaches, e.g. to allocate the investment costs in terms of a split into load and non-load related elements (as discussed elsewhere in this report for overlays; see e.g. section 7.3), and the allocation of load-related costs on the basis of ESALs, see e.g. Ritlett et al. (1989)

The club approach takes this thinking two steps further[33]. The first step is to ask, whose values should determine what is meant by 'equitable'. In terms of the club approach it is the values of the members of the club that count, i.e. road users should decide how costs should be allocated. The second step is then to try to understand how road users would actually themselves allocate costs, if they somehow were able to 'sit down' and negotiate how to share the costs. A theoretical approach to this is by way of game theory. Game theory has been applied in a number of studies to understand how e.g. railway companies would agree to pay for the use of joint facilities (a railway station and related infrastructure, Brown (1985)), and also the structure of fees for making use of uncongested runways in aviation (Littlechild and Thompson (1977)). The same thinking has recently also come to be applied in order to allocate road costs in Germany, albeit in a more schematic way; see further below.

The reasoning can be illustrated with reference to the work by Littlechild and Thompson[34]. Their approach involves game theory to allocate common costs to different types of movements/aircraft[35]. Three basic principles are used as the starting point for deriving the pricing policy. These are that (i) prices should provide for cost-recovery, (ii) that each type of aircraft should at least cover its own marginal cost associated with a landing, and (iii) that no type of aircraft should pay more than the stand-alone cost for providing a runway for landing of that type of aircraft. These principles are assumed to be acceptable and supported by the members of the club.

In determining prices meeting these three requirements, Littlechild and Thompson establish the following rule for how to allocate the – fixed – construction costs of a runway: "the cost of the runway required by the smallest aircraft is divided equally among the movements of all aircraft types, the cost of the next increment of runway is divided among all movements, except those of the smallest aircraft type, and so on until the cost of the last increment of runway capacity is divided only among the movements of the largest aircraft"[36]. Prices set according to this rule were found to not be very dissimilar from the ones actually imposed at the particular airport which was examined in their study (Birmingham, UK); these increase with the size of the aircraft measured in terms of maximum take-off weight (MTOW)[37].

---

[33] In economic theoretical terms, a club is defined as a voluntary group deriving mutual benefit from sharing one or more of the following: production costs, the members' characteristics, or a good characterised by excludable benefits. Clubs do not exist primarily in order to generate a return on the capital invested; they serve the wider ambition of furthering the interests of the members of the club. While clubs can operate on commercial terms, they are not necessarily commercial entities. The beauty of the club approach is that if a public good can be operated in such a way, it obviates the need to be concerned with the formulation of the objectives for the operations of the public good. In a club, the members decide directly through the decision-making mechanism what they want to achieve and how; see Sandler et al. (1980) and Cornes et al. (1986).

[34] Littlechild & Thompson (1977).

[35] To find a solution both the Shapley value and variations of the nucleolus were used; for an explanation, see e.g. Sandler et al. (1980).

[36] Littlechild & Thompson (1977), p. 201.

[37] The main determinants of the level of the charge are manoeuvrability, the length of the runway required and the footprint pressure of the wheels. The maximum take-off weight (MTOW) of an aircraft can be seen as a proxy for these three variables.

The above presentation illustrates the similarities and the differences between the club and the equity approach, on the one hand, and the approach based on marginal cost-pricing, on the other hand. In the club approach, as e.g. developed by Littlechild and Thompson, every user should bear his own cost, essentially the marginal cost of use. In cost allocations exercises, this cost is normally calculated by making use of assumptions implying constant variable cost of maintenance. This approach is essentially the same as the indirect approach as presented above, when the assumptions underlying Newbery's theorem may be invoked.

The difference between the two approaches is that the marginal cost approach does not pay any immediate attention to the fixed costs. In terms of this approach, the fixed costs are related to capacity and its scarcity value is reflected in the cost of congestion, a subject matter not covered in this Survey. In cost allocation exercises, on the other hand, fixed costs and their allocation play a central role. As illustrated above, it is thus normally assumed that each class of vehicle user should bear the incremental cost associated with that class, and these costs common to a class are to be distributed equally amongst the users in that class. Remaining costs are to be borne by all users based on a neutral criterion.

It is against this background that space is given in this Survey to briefly present cost allocation exercises based on the equity or club approach. These are the most common form of road cost studies, and it may be claimed that they produce results which could be seen as being of interest also to the marginal cost-approach.

5.2 British studies

The most recent British cost allocation study has been performed by NERA on behalf of the Department of Transport[38]. It follows the approach of earlier British cost allocation studies. The aim is to fully allocate the total annual infrastructure costs of the road system. One of the basic assumptions of the approach is that once a road has been constructed it 'lasts for ever'. As a consequence, the annual capital cost may be represented by the interest cost of the capital originally invested, plus all costs required to preserve and operate the road network. A second assumption underlying the methodology is that traffic dependent maintenance costs are determined by some rule of optimality. Maintenance costs may hence not correspond to actual expenditures incurred by the road authority.

The allocation methodology used may be interpreted in terms of the 'club approach'. The allocation principles are summarised in the Annex (at the end of this Survey), showing that the following four measures of road use are employed: (i) PCU-km; (ii) Average vehicle weight-km; (iii) GVM-km; and (iv) ESAL-km. In addition, v-km are used to allocate costs which are common to all vehicles.

The allocation of the capital charges may be interpreted to say the 85% of investment costs are due to capacity needs in general, whilst 15% are required to meet the needs of heavy vehicles in particular. Capacity costs (and congestion costs, the dual of capacity costs), may be seen to be allocated in terms of PCUs, which reflect the capacity needs in terms of different types of vehicles; see e.g. Newbery (1988a). The measure, GVM-km, indicate the difference in incremental investment cost incurred by vehicles with different GVM.

[38] Department of Transport (2000).

As concerns maintenance costs, the three first categories may be seen as caused mainly by the damage of heavy vehicles, whilst surface dressings are required more on account of the wear of vehicles. Drainage is a common cost to all, and therefore cannot be allocated according to a vehicle specific logic; it is therefore allocated on an equal basis in terms of v-km, etc. Implicitly, the allocation principles used in the UK could be seen to be in conformity with the principles proposed by Littlechild and Thompson.

In addition, the cost estimates provided by the NERA study may be used to estimate the marginal costs for different types of vehicles, and also different types of roads. As mentioned, for this to be appropriate, maintenance has to be seen as determined based on a condition-responsive approach (as apparently done by the study). To obtain marginal costs, it must also be assumed that the weathering effect is equal to zero (as implicitly assumed in the study), or alternatively, additional assumptions about this effect must be introduced.

In further research commissioned by the Department of Transport, the Institute of Transport Studies (ITS) of Leeds University have used the NERA calculations to determine short-run marginal costs with respect to, *inter alia,* infrastructure use[39]. Although not stated, the implicit assumptions made use of in the ITS study are the ones mentioned in the previous para. Their marginal cost estimates reflect all routine and periodic maintenance costs which could be seen as variable (given no weathering effect)[40]. Estimates were obtained for 5 different types of vehicles (cars, light delivery vehicles, rigid heavy trucks, articulated heavy trucks and buses) and 3 different road classes. On average, for vehicles and roads, the marginal infrastructure cost was estimated to vary between 0.42 to 0.54 pence per v-km. Estimated marginal costs ranged between 0.05–0.07 pence per v-km for cars to 7.55–9.82 pence for an articulated heavy truck; all estimates are in 1998 prices.

5.3 US federal studies

The US federal authorities have a long tradition of carrying out cost allocation studies based on the equity approach[41]. Major studies were carried out in 1982 and 1997 (with an update in 2000)[42]. The focus of the federal studies has been on allocating the federal expenditures on roads (from the Federal Highway Trust Fund), which implies that only part of road expenditures have been considered. In the US, the states and local authorities are, in general, the principal road authorities, and spend the major part of monies of the road network, although federal contributions make up a significant portion. The emphasis of the federal road allocation studies have hence been on trying to ensure that federal taxes levied on different categories of road users correspond to the expenditures on these road users. However, the most recent study, in 1997, also took steps to identify the marginal social costs of road use as well as the total expenditures on roads. In recent federal cost allocation studies there is hence information also of

[39] See Sansom et al. (2002).

[40] A strange aspect of their marginal cost calculations is that they apparently include also drainage costs, which cannot be seen to be caused by a marginal vehicle.

[41] A number of the states have also carried out their own cost allocation studies. It is reported that 32 states had carried out such studies by the early 1990s; see Nix (2000).

[42] See US Department of Transportation (1997) and (2000).

relevance to marginal infrastructure cost calculations, although the data are not immediately available in the documentation that has been published (on the net).

The federal cost allocation method is in principle based on the club approach (for federal expenditure). The methodology used is referred to as the 'cost occasioned approach'. As for UK studies, the ‘occasioned’ cost is either a variable cost, then treated as a marginal cost, or a cost associated with a particular type of vehicle, e.g. the cost of additional pavement bearing capacity required by heavy vehicles. A difference in comparison with the UK studies is that weathering is considered explicitly. Another difference is that the US cost allocation study attempts to allocate all expenditures of a particular year to the road users, including capital expenditures. As mentioned, capital costs in the UK methodology is reflected in the form of depreciation. The US federal methodology is based on detailed research and data. It is understood, however, that unlike the recent UK study, it does not make use of data on the costs of maintenance which have been generated based on some assumption of optimality.

In very broad terms, investment expenditures on additional capacity are divided into 'base facility costs', i.e. costs required by all vehicles, which are allocated on the basis of PCE-km, and incremental costs. PCE is similar to PCU, used in the UK, but is a more refined measure of the difference in demand for capacity by different vehicles. The incremental investment expenditures on new capacity are allocated, in principle, in terms of ESAL-km[43]. Similar principles are used to allocate the costs assumed to vary with traffic for periodic maintenance and some routine maintenance activities (resurfacings). All costs which are seen as residual or common to all vehicles (including weathering) are allocated on a v-km basis. In the federal system, bridge costs and so called system enhancement costs (safety improvements, etc.) are also allocated. 20 different types of vehicles are identified.

The (short-run) marginal cost calculations performed as part of the Federal Highway Cost Allocation study includes the following elements: Air pollution, noise, congestion, crash costs and waste disposal costs. Marginal pavement costs are also included, and are defined as "the contribution of a mile of travel by different vehicles to pavement deterioration and the costs of repairing the damage", no further details are provided in the main report as to how these marginal pavement costs have actually been computed[44].

5.4 The EU study

In 1997, the EU Commission initiated a study entitled "*Infrastructure Capital, Maintenance and Road Damage Costs for Different Heavy Goods Vehicles in the EU*". It was carried out by a consortium of consultants, and eventually published in Link et al. (1999). It contains a review of the then state of the art with respect to road costing and cost allocation for roads in Europe. The study, in addition, attempts to prepare new estimates of total costs of the road networks for the EU member states plus Switzerland, including marginal costs, comprising infrastructure and congestion costs.

[43] Account is also taken of additional grading and drainage needs due to weight and width of vehicles.

[44] Reference is made to an appendix E, which has not been available.

The data used were obtained from member states (and Switzerland), and the methods applied were also essentially those of the member states, where such methodology exists. The study was performed during a short time period, and its documentation is incomplete. In addition to total costs and allocated costs, the study presents estimates of marginal infrastructure costs for some of the member states, including Denmark, Finland, France, Germany, Ireland, Portugal (only heavy goods vehicles), Spain, Sweden, Switzerland and the UK. Estimates for 6 different vehicles for the entire domestic networks as well as the motorway networks are presented for most of these countries. The estimates are in 1994 ECU, having been converted to a common price level by using PPPs.

The description of the methodology used is sparse. It is only stated that maintenance expenditures for pavement and resurfacing, and parts of the reconstruction expenditures are included. It is presumed here that the allocation principle has been decided on with reference to the method used in each particular country. Apparently no consideration has been taken of weathering.

5.5 The German study

Germany has a long history of road cost and cost allocation studies. The latest was prompted by the decision of the German federal government in 2001 to replace the current Eurovignette for use by trucks of the motorway system by a weight-distance charge, to be introduced in August 2003. The new charging system will differentiate between two types of trucks, and charges will vary depending to the EURO-standard for vehicle emissions.

In order to determine the level of the km-charge, and to be in compliance with current EU-legislation, the German transport ministry commissioned a cost allocation study[45], with the aim of fully allocating costs of the German motorway system to different types of vehicles. The German study is one of few studies in which explicit reference is made to game theory in order to provide a justification for the methods used for allocating costs[46]. It makes a distinction between three categories of costs based on (i) causality (in principle marginal cost), (ii) specificity (incremental costs above a base facility), and (iii) fairness, i.e. the principle for allocating common or residual costs. Costs have been calculated with respect to all federal roads, and not only the motorways of the country.

Whilst similar in its approach to the UK and US Federal studies, the German study has some distinct features. It contains an extensive analysis of capital costs, but the principles applied in 'measuring' these costs are different from the UK and US studies. Capital costs include depreciation and interest. Allocation principles of these capital costs are simplified in comparison with the UK and US studies, in that the study differentiates between 6 types of vehicles only, and the main issue is to identify the costs associated with trucks with a GVM of 12 ton plus. Principles of allocation are similar as in the other studies, but the costs of a (short-run) marginal cost nature include only resurfacings and other variable routine maintenance costs. The resurfacing costs are allocated in terms of ESALs (using the 4th power rule). Other variable routine maintenance costs related to the condition of the road surface are allocated to heavy goods vehicles (12 tonnes and

[45] See IWW and Prognos (2002) and Rothengatter (2002).

[46] Another study making reference to game theory is the one by Ghaeli et al. (2000). The source provides inadequate information to enable a presentation here; reference is made to this study below in the section on econometric studies.

plus) and on the basis of axle-km. Residual operations and routine maintenance costs are seen as common costs to all vehicles and allocated on the basis of v-km. Overlay and reconstruction costs of the base are treated as an investment and, hence, as an incremental cost – to be borne by heavy goods vehicles (12 ton plus) based on v-km – and are therefore not reflected in marginal costs. Calculated marginal costs refer to averages for the entire motorway network. The German approach does apparently not consider weathering. It is unclear how the costs of maintenance actions have been determined. The study refers to data obtained from the Ministry of Transport but does not say how the maintenance budget has been prepared, e.g. whether a condition-responsive approach is used or not.

From the documentation available, it is not possible to extract information about marginal infrastructure costs, as these costs – implicitly – are defined in the German approach.

5.6 Other countries

Full cost allocation studies have been carried out in a number of other countries. Officially sanctioned studies in other European countries may be found in e.g. Austria, Denmark, Finland and Switzerland[47]. Amongst overseas countries the following should be mentioned: Australia, New Zealand, South Africa and Namibia[48]. The studies performed in these countries are of a similar vein as the studies already referred to here, but each will normally have its particular characteristic. Mention should also be made of studies carried out under the auspices of the World Bank, normally making use of the RUC 30 model; these studies may essentially be seen as cost allocation exercises although explicit reference is made to marginal costs[49]. Finally, there are examples of research studies, not sponsored by the government, or a road authority, of the country/state/province concerned[50].

[47] For a review, see Link et al. (1999). The Swiss introduced a km-based charge on vehicles of GVM of 3.5 tons and plus as from 1 January 2001. The km-fee was determined by reference to three cost components, viz. (i) external costs, (ii) compensation for income for the state on account of the replacement of the previous flat fee by the km-charge, and (iii) to ensure a balance of revenues and expenditures in the road sector. The sum of these components was then divided by the estimated ton-km driven p.a. on Swiss roads to obtain a cost per ton-km. The Swiss study is hence of limited interest from a marginal cost point of view.

[48] For Australia see NRTC (2000), for Namibia see MWTC (1997), for Southern Africa see Joint Task Team (1995), for South Africa see Department of Transport (1989), and for New Zealand, see Ministry of Transport (1995).

[49] For an example, see Archando-Callao (2000).

[50] See e.g. Ghaeli et al. (2000) and a study by Gaudry, Mallet and Marullo, referred to in Nix (2000).

6 Econometric Approaches

6.1 Introduction

Estimates of marginal costs obtained by way of the direct and the indirect approaches are at times referred to as estimates based on the engineering approach. An alternative sometimes advanced by economists, e.g. in the UNITE project, is the econometric approach[51]. This latter approach, in principle, envisages a relationship of the following nature (which can be justified with reference to microeconomic theory):

$$TC = f(TI, W, S, K, C) \qquad (6.1)$$

where TC is either the total cost of building and preserving a road over its lifetime, or the total cost of preserving it over its lifetime, TI is the output as measured by road use variables (v-km, GVM-km, ESAL-km, etc.), W refers to the geometry, e.g. the width, S is the initial structural condition of the pavement (or similar), K is a measure of climatic impact, and X is a vector of other variables explaining costs.

By first estimating the parameters of such a relationship with econometric methods and then differentiating the estimated relationship, it is possible to obtain an estimate of the marginal cost with respect to increased road use. If the dependent variable is total cost, some estimate of long-term marginal cost is obtained; if the dependent variable excludes the original investment costs, then a measure of short-run marginal cost is obtained.

To actually apply this approach is not easy as demonstrated by that there are few examples of econometric studies available today. One issue is that of generating a data set given that observations reflecting the past are to be made use of. A relationship like (6.1) will have to be estimated for a number of road sections of a road network. For each such section observations of costs will have to made over a long period of time in order to generate a longitudinal data set reflecting the actual costs associated with that road. This means that the road will have to have been through a number of maintenance actions in order to be able to obtain an estimate of the cost associated with that road. Similarly, observations for the independent variables will have to be made for the same time period. Data of that nature are not normally available, although the management information systems (MISs) now being made use of by road authorities will in time be able to generate at least in part data of this kind.

The reason for suggesting that even the data bases of modern MISs may be inadequate is the following. Firstly they may not refer to an adequately small segment level to be of interest to an econometric analysis. Secondly, whilst modern MISs provide plenty of information about roads, including about interventions in the road, data on road use tend to be limited. MISs typically provide data on e.g. AADT for different types of vehicles, including heavy vehicles, observations on such road use variables as GVM or ESALs, which would be critical to a successful application of the econometric approach, are not normally included..

[51] See e.g. Link and Lindberg (2000) and van den Bossche et al. (2001).

Assuming that data would be available, a second issue arises, and that is to actually disentangle the contribution to overall marginal cost of different road-use variables. Road-use variables tend to vary in the same fashion, so that higher v-km is associated with higher GVM-km and ESAL-km. Higher ESALs also tend to vary with the structural condition of the pavement, reflecting the traditional approach to the design of a road, etc[52].

A third issue related to the econometric method is that it does not account for costs associated with road damage externalities. Observations on costs only relate to maintenance, reconstruction, etc., and do not reflect the impact on other road users.

6.2 The Link and related studies

The Link study (Link (2002)) is one of few in which use is made of estimates of TC by way of longitudinal observations[53]. The data used pertain to twenty years (1980–1999) for different segments of the former West German motorway network. The costs include 'major renewal' costs, presumably reflecting only the costs of overlays. The only measure of road use that apparently was available to the study was accumulated v-km over the time period for different types of vehicles. Additional data made use of include climatic data, the age of the segment of the road studied, and expenditures on each studied segment before 1980.

The trans-log functional form was employed in this study. Only part of the variation in the dependent variable was explained by the independent variables. The best fit was obtained when traffic was represented by way of the ratio between v-km for cars and trucks. The usefulness of the study is limited by the fact that the main explanatory variable is a ratio and also does not include ESALs. The main factor explaining renewal costs is likely to be a measure of road use in terms of this latter variable.

An issue related to this study is how to interpret an estimate of the marginal cost obtained from the model as estimated in this particular study. The TC is apparently a summed variable with respect to all renewal actions during the studied 20 year period. Such a variable will not only vary on account of the level of road use (ESALs), but also with respect to where in the maintenance cycle an observation is made. During the observed period of twenty years, a particular segment may thus have been through one or two overlays, cet par. One or two overlays will yield very different estimates of the TC variable. It is noted that Link's model does include a variable reflecting expenditures before 1980, but that variable may reflect primarily the overall age of the road.

[52] This may in effect not necessarily be the case. However, the fact that detailed data on road use is not normally available for individual roads, necessitate the estimation of such data based on aggregate data. Such an approach will result in a tendency for road use data to vary in the same way.

[53] Whilst longitudinal data are used, the observations for different years with respect to one particular segment are summed over the entire observation period. Individual years are hence not used, which means that important information is lost. Longitudinal data in other econometric studies are aggregated in a similar way. However, in the study by Johansson and Nilsson (2001) for railway track maintenance costs, data on costs for each year are used.

Two other similar econometric studies have been performed with data from Austria and Switzerland[54]. Both rely on more comprehensive data on maintenance costs for various segments of the motorway and road network (Switzerland only) than the German study. However, the data refer to short time periods, 1985 to 1998 for Swiss roads, 1997 to 2000 for Swiss motorways and 1990 to 2000 for Austrian motorways. The definition of costs (TC) appears unclear, but it does include both routine and periodic maintenance costs. Limited information on road use was available, and the road use data employed in these studies have essentially been measured based on cordon counts at various locations, in order to estimate v-km for different types of vehicles, and by making use of separate studies related to axles and GVM. Since these studies are not based on actual observations of road use for the road segments for which cost data are available, the data are likely to be suffering from a lack of variability as concerns different types of vehicles. No data on ESALs were available.

The functional form used in these two studies was the log-linear. The dependent variable was maintenance costs for the different segments summed over the period for which data were available. It proved not possible to disentangle the effects of the road use variables for different types of vehicles. The 'best-fit' model for Swiss data explained costs in terms of accumulated GVM-km; similarly the 'best' Austrian model made use of accumulated v-km for the studied period.

The German study yielded estimates of marginal costs for trucks varying between 0.05 and 2.70 € cents/v-km, the Austrian study estimated the marginal cost of trucks to 2.17 € cents per v-km and of cars to 0.07 € cents per v-km, and the Swiss study yielded values between 3.62 and 5.17 € cents per v-km for trucks and 0.42 to 0.40 € cents per v-km for cars. The data problems associated with these three studies suggest that the indicated estimates of marginal costs are of limited value.

6.3 Studies by Li et al. and Martin

Two other studies make use of an approach similar to Link's but in a simplified format, and also appear to be able to handle the 'age' problem in a more explicit way[55]. The study carried out by Li et al. (2001) makes use of expenditures on rehabilitation and periodic maintenance in order to determine the dependent variable. This variable has been estimated based on the total expenditures during a cycle, i.e. the period from one reconstruction to the next, starting for the year after the completion of a reconstruction of the road. Further details are, however, not available. Data for a number of road segments are being made use of and for different types of pavements. The explanatory variables include, *inter alia*, change in IRI during one life cycle, climatic and regional features, cumulative ESALs, pavement age at the time of the rehabilitation, and the thickness of the pavement.

The estimated functions were also used by Li et al. to estimate the marginal cost with respect to an ESAL. That marginal cost includes both rehabilitation and periodic maintenance costs. The estimated relationships for how both total reconstruction and maintenance costs and marginal costs vary with traffic loads

[54] See Link et al. (2002).

[55] A further econometric study is Ozbay et al. (2001), which refers to Small et al. (1989). The Ozbay study relies on vehicles per day as the explanatory variable of resurfacing costs, but provides inadequate information to allow for an interpretation of the approach used.

(ESALs), was in the final step used to determine the non-load and load-related shares of pavement rehabilitation (i.e. reconstruction plus periodic maintenance) expenditures. Data for the state of Indiana in the US and for a flexible pavement indicated a load-related share of 42%. The other 58% of the costs would be explained by such factors as climate and time[56].

In a similar study carried out by Martin (1994) but with Australian data, the dependent variable reflected average pavement related routine and periodic maintenance costs for various road segments. These averages were computed based on observations over a number of years, although in general the time span was quite short, 5 years. Different road use variables were tried, including cumulative ESALs-km and cumulative GVM-km. Another explanatory variable used was pavement age. The findings for data from different states in Australia and for different types of roads indicated no uniformity as to which road use variable that would best explain the variability in the data. Age, measured as the number of years since the construction or reconstruction or rehabilitation, was only significant in some of the estimated relationships. Martin concluded that on average 53% of total pavement related maintenance expenditures could be explained by road use (measured in terms of GVM per lane and year).

6.4 The Ontario, Canada, studies

Two econometric studies based on data from Ontario indicate how the data problem associated with the econometric method can – if not overcome – then at least be reduced; see Hajek et al. (1999) and Ghaeli et al. (2000). In these two studies, the data were synthesised. The approach used to obtain the set of observations started by making assumptions about the traffic load in ESALs on different roads and the growth in this traffic over time. Costs for the studied roads were then generated in the first of these two studies by applying the rules set out in the manuals for construction and maintenance used by the road authority for when to undertake periodic road maintenance, reconstruction and expansion to provide additional capacity. The costs over a sixty year period were identified and then discounted and converted to an annual cost. In addition unit costs for various types of road works were made use of.

Four different sets of cost data were generated in this way relating to new road capacity and existing roads, on the one hand, and northern and southern Ontario, on the other. The following relationships were estimated for old roads (only maintenance costs) and new roads (also including construction costs in the annualised costs):

$$EUAC = a + b\ln(ESAL)^2 + cN + e$$

where EUAC is the equivalent annual cost (the annuity) per lane and N is a dummy for southern and northern Ontario. Good statistical fits were obtained, and the estimated functions made it possible to determine a short-run marginal as well as a long-run marginal cost with respect to an ESAL. It is, however, unclear if the marginal cost is to be seen as determined for a an average age in terms of e.g. an overlay cycle, for roads carrying a certain level of ESALs, or if the estimate

[56] In a similar study, using a similar approach, Li et al. (2002) examined the load and non-load related shares with respect to routine maintenance. The load share for flexible pavements was found to be 26 % for data from the state of Indiana, USA.

should be seen as representing a marginal cost at a certain time of the cycle, e.g. after a road has been constructed or just before or after an overlay.

This points to a common question related to the econometric approach. The total cost (either including or excluding the original investment cost) is often a present value or based on a present value, and the assumption as to which year it pertains will affect its size. A relationship such as (6.1) must therefore contain an explanatory variable to reflect the age of the road within a cycle (e.g. an overlay cycle or reconstruction cycle). It is likely to be a challenge to identify an appropriate variable in this regard; see also the comment made above with regard to the Link-model. In the second study, the data were generated with the PMS used by the Ontario road authority. In this study cost functions were only estimated for new roads.

7 A Review of Some Common Issues

7.1 Introduction

Two issues that all the methods reviewed face as concerns the estimation of marginal costs are the appropriateness of the rule of the fourth power and the contribution of weathering and other non-traffic effects to road deterioration.

7.2 The damage caused by heavy vehicles

The main issue related to relationship (2.1) is not its functional form but the size of the coefficient ω, i.e. whether it should be another value than the frequently used value of four, or if different values should apply to roads with specific pavements and nature of traffic. The power of four rule was established based on the AASHO-tests carried out in the 1950s[57], and since that time both the way roads are built and maintained and the heavy vehicles (trucks) have changed and developed.

The importance of this debate to the estimation of marginal infrastructure cost is obvious. Assume that the coefficient ω is 3 rather than 4. The number of ESALs of a particular vehicle will then be reduced in half in comparison with a vehicle with a standard axle. Since ESALs is the main factor driving marginal costs, the ratio between the marginal infrastructure costs associated with heavy and light vehicles will be reduced significantly.

There are also questions with regard to how the original research generating the fourth power law was carried out. The Brooking Institution study (Small et al. (1989)) re-estimated the relationship (2.1) using the original road test data, and found exponents to be about 3, or generally lower than the original AASHTO estimate of 4.

Subsequent research in Europe indicates that the coefficient will vary depending on the structural condition of and type of pavement. The coefficient has thus been found to range from 2.5 to 5.5 in a Nordic project, whilst an OECD study reports an interval between 2 and 9. The NERA study on cost allocation summarises the situation as follows[58]:

> ”There is no convincing evidence that the 4th power of axle weight is not the most appropriate general rule to allocate structural pavement maintenance costs. However, it does appear that different power rules may be appropriate for different types of roads, with a lower power function for motorways and a higher one for minor roads.”

NERA has carried out an analysis of the importance of actually using more detailed information on the ω coefficient, making use of different values for different types of roads. From the point of view of a cost allocation exercise, the more disaggregated approach proved not to have large implications. It is to be recalled that one of the main purposes of cost allocation is to establish the fairness of the road taxation system, on average, and not the fairness with respect to individual roads. This also means that if marginal costs are to be determined as an

[57] Highway Research Board (1961).

[58] OECD (1988) and Department of Transport (2000), p.51.

average for a road network detailed information of this coefficient may not be critical. If, on the other hand, marginal infrastructure costs are to be determined for each road, the choice of the value on the coefficient will play an important role.

When the 4th power law was originally proposed most heavy vehicles had conventional steel suspensions. More recently, in particular heavy trucks have come to make use of so-called ’road-friendly suspensions’, i.e. air suspensions. It has sometimes been claimed that such suspensions would lead to less road damage. The DEVINE experiments carried out by OECD reported that pavement surface profiles deteriorate more rapidly under loading applied by a steel suspension compared with loading applied by a ’road-friendly suspension’[59]. Other studies have shed different light on this question. At least one study (Martin (2000)), referring to recent research and Australian conditions, conclude that there is no clear evidence that air suspensions cause less pavement damage.

Another issue concerns the replacement of dual tyres with wide single tyres, which reduce operating costs, but apparently can increase the impact on the road by 50 to 100%. Research is currently ongoing in Europe via COST to analyse in further depth the cost and benefits of these tyres. Apparently no work has as yet been done on how to account for these vehicles when calculating ESALs[60].

7.3 Damage and wear caused by climate and other non-traffic sources

The importance of the ‘weathering’ effect has been indicated before; as for the size of the coefficient ω in relationship (2.1), it has significant implications for the estimation of marginal costs for different types of vehicles. The Newbery approach implies that the marginal cost of an overlay is, normally, smaller than the average cost; the Small et al. (1989) model suggest that the reverse holds, although their approach starts from very similar premises as Newbery’s.

The assumptions underlying the Newbery approach predominate in empirical work of marginal costs of road use, and are also reflected in the HDM 4 model. They imply that road user charges set to equal the marginal cost of road use will not recover maintenance and reconstruction costs. Such charges will thus recover neither costs on account of weathering nor other routine maintenance costs which are not related to the pavement (e.g. drainage and vegetation control)[61].

The assumptions as to the size of the weathering effect vary between studies. In his studies for the UK, Newbery assumes weathering to contribute 60%. In the study undertaken by Vitaliano and Held (1989) for roads in New York State, the weathering factor is set at 50%. A recent Swedish study is, as noted, based on a zero weathering effect, reflecting findings from empirical studies. Also, as concerns cost allocation studies, different assumptions may be found. The US cost allocation studies are based on very detailed assumptions about the contribution of weathering to the cost associated with different routine and periodic maintenance

[59] See Department of Transport (2000).
[60] See Department of Transport (2000).
[61] However, as shown by Newbery (1989) as well as Small et al. (1989), these costs may be recovered, under certain circumstances, if congestion costs are also reflected in road user charges.

activities. The UK and German cost allocation studies, on the other hand, do not consider weathering[62].

Some of the studies carried out to actually determine variable and fixed maintenance costs or load and non-load costs have already been referred to above in the context of the econometric studies. Econometric techniques have thus sometimes been employed for this purpose, by regressing total maintenance costs on measures of road use. Another approach has been made use of by Martin (1994) – also involving econometric techniques – to estimate the share of routine and periodic pavement related maintenance costs explained by heavy vehicles. He examined and compared changes in roughness (measured as IRI) at different locations on a road, in principle comparing roughness estimated at a location where the wheels pass with locations where they rarely pass. He then adjusted his data to account for the fact that distress (in the form of roughness) in the wheel path of a vehicle is transmitted to other areas of the road not so frequently used, and obtained for a sample of roads data on the ratio between load related wear (measured as IRI) and non-related wear. This dependent variable was regressed on explanatory factors such as road use, the structural condition of the pavement, and a measure of the climatic impact. He found that either cumulative ESALs, or cumulative GVM, per lane and year, best explained the variation in the dependent variable. His conclusion was that for the Australian roads examined, 50% of routine and periodic pavement related maintenance costs could be attributed to heavy vehicles on the basis of GVM-km. The study was subsequently updated with additional data, and it was then found that ESALs was a preferred explanatory variable to GVM (Rosalin and Martin (1999)).

In a similar study, Martin (2001) instead of observed data made use of simulated data obtained by way of the HDM III model calibrated for Australian conditions. Again costs for pavement related maintenance were assumed to be directly related to change in roughness, and hence costs attributed to heavy vehicles were estimated as the portion of change in the measure of roughness caused by heavy vehicles in relation to the total change in roughness. As for the other studies, about 50% of deterioration was found to be explained by traffic; road use was in this study measured by cumulative ESALs.

[62] Other countries also assume that part of the costs of routine and periodic maintenance may be seen as fixed (independent of road use) and therefore should not be allocated on the basis of e.g. loads; see Link et al. (1999).

8 Concluding Words

This Survey seems to lead to the following conclusions:

1. Marginal infrastructure costs are heavily influenced by the damaging effects in the form of various types of distresses caused principally by the loading of traffic. Loading is measured by way of ESALs, and normally the fourth power law is made use of. This law is also used commonly in calculations of marginal costs, but its relevance is somewhat in question. This is an important issue in that the size of the power coefficient may significantly affect the ratio of the marginal infrastructure cost for heavy and light vehicles. However, this is a common issue to all approaches used today to measure marginal infrastructure costs.

2. UNITE documentation makes reference to econometric approaches in order to measure marginal infrastructure costs, but there are few applications of this approach. The reason is likely to be that it is difficult to generate the required data – longitudinal data over a long period of time for 'segments' of roads will likely be required – and it seems unlikely that road authorities will have such data in their MISs. Also, it will be difficult to disentangle the effects on maintenance of various measures of road use (v-km, GVM-km, ESAL-km, etc.), as they tend to vary in the same way. Besides, the econometric approach normally (albeit not necessarily) makes use of historical data. Presumably, estimates of marginal costs should be based on current or expected future cost data, as marginal cost based pricing is supposed to guide resource use today and in the future. Mention should also be made of that the econometric approach does not account for costs for road damage externalities.

3. A distinction is sometimes made in the literature between the cost allocation and marginal cost approaches when determining costs related to road use, seemingly suggesting that the two approaches are incompatible[63]. We do not think that the distinction is necessarily substantive. Cost allocation exercises are generally based on the notion of equity; a theoretical basis for equity is the club approach, one central aspect of which is that a marginal cost should be borne by the user giving rise to this cost. Since cost allocation exercises normally are based on the club approach, they should be of interest to a review of estimates of marginal infrastructure costs.

4. Indeed, it is argued here that both what is called the indirect approach and the full cost allocation approach (when based on the club approach) basically rely on the properties of Newbery's 'fundamental theorem'. Our understanding of this theorem is that it allows for ignoring the road damage externality effect, and for deriving marginal costs from estimates of average costs. Our understanding is furthermore that Newbery's theorem is primarily of importance to overlays (the cost of which makes up a very large – and apparently growing – share of maintenance costs), but that it may be applied also to other maintenance costs which arise as a consequence of road use, e.g. certain routine maintenance costs, and presumably also reconstruction costs. It is

[63] See e.g. Transportation Research Board (1996).

noted that this possible extension of the Newbery approach apparently has never been developed in full.

5. This notwithstanding there are other issues related to the Newbery approach, as different researchers and analysts use different methodologies, which are related but result in different conclusions. Newbery's conclusion is that marginal costs with regard to overlays may be calculated as the average cost per ESAL corrected – and reduced – to account for the weathering effect. The marginal cost will therefore be lower than the average cost, and optimal road user charges will therefore not recover road maintenance costs. Small et al., using a similar modelling approach as Newbery, find that the marginal cost of an overlay with respect to an ESAL is higher than the average cost. Lindberg uses a different – more general – relationship than Newbery and Small et al. between the measure of a road's condition and road use measured in terms of ESALs. In terms of his findings, therefore, the average cost has to be corrected by a further term, the 'deterioration elasticity'. These differences in modelling approach are not trivial when it comes to the calculation of marginal costs. They likely have ramification also with respect to other types of maintenance and reconstruction costs, although no analysis thereof is available.

6. An important assumption underlying 'the fundamental theorem' is that maintenance actions are rule-based or condition responsive. Normally it is assumed that once roughness exceeds a certain value then the road will be overlaid. Such principles are also followed by road authorities, albeit perhaps more in theory and planning than in reality. This raises the question: If the planning framework is assumed to be relevant, and if, in addition, prices should be tailored to marginal costs so that they in effect influence present and future road use, then why not use the planning models (pavement management systems (PMS)) used by road authorities to determine marginal infrastructure costs? The idea is not new, but has rarely (if ever?) been applied. It has been demonstrated here that it may be used, as illustrated with the application of the HDM 4 model for a specific road.

7. The implications of using a PMS (such as HDM 4) are the following. It has to be recognised that these models embody certain assumptions reflecting the effect of weathering and the nature of the relationship between road use and change in road condition. It is our understanding that with respect to overlays, the HDM 4 model is close to the assumptions used in the Newbery approach (with implications as concerns weathering and the deterioration elasticity). On the other hand, using a PMS obviates the need for relying on a partial approach as characterises the indirect method as well as the econometric approach. As demonstrated by calculations presented in this Survey, the more comprehensive approach to wear and damage reflected in a PMS such as HDM 4 could well lead to the conclusion that the road damage externality component ought not to be ignored. There are two further aspects to be noted. Firstly by using a PMS approach the marginal cost associated with a particular type of vehicle is calculated directly; it is not necessary to identify separately the marginal cost for each one of the road use variables, (assuming that ESALs do not capture all effects of road use). Secondly, it directly takes into account all types of maintenance actions, i.e. routine, periodic and reconstruct-

tion, etc., resulting from road use. Other advantages are that the approach relies on unit costs for various inputs into different types of works activities, and can be tailored – by calibration – to reflect the conditions in a particular country and part of a country, including the particular nature of roads as concerns pavement, base and sub-base. In addition, by using a PMS such as the HDM 4 it will be possible to determine marginal costs at different levels of aggregation, e.g. either for a specific road or as an average for a road network. Newbery's approach can only be applied for a road network.

8. A number of estimates of marginal costs, expressed sometimes in terms of ESALs, sometimes per km for different types of vehicles, have been presented. No effort has been made to compare the results in view of the differences in methodologies used, the nature and quality of the data employed, and that many empirical results refer to situations which may not at all be relevant to the conditions where marginal cost based pricing is to be applied. This Survey demonstrates that it is possible to obtain estimates of marginal road use costs, although the quality will vary; however before making use of such estimates, the specifications that should be applied with respect to estimates of marginal infrastructure costs must be formulated.

9 References

Archondo-Callao, Rodrigo S. (2000): **Road User Charges Model**, Africa Transport Technical Note, Note No. 28, Sub-Saharan Africa Transport Policy Program, The World Bank, Washington, D.C.

Brown, Stephen M. (1985): **Cost-sharing Clubs and the Private Provision of Transportation Infrastructure,** Ph.D. thesis, Washington University, St. Louis, Mo.

Commission of the European Communities (2001): **White Paper: European transport policy for 2010: time to decide,** COM(2001) 370, Brussels, 12/09/2001.

Cornes, Richard and Sandler, Todd. (1986): **The Theory of Externalities, Public Goods, and Club Goods,** Cambridge University Press, Cambridge UK.

Department of Transport (1989): **Final Report on the Reassessment of Road Infrastructure Cost Recovery,** Prepared by van Wyk & Louw Inc., Pretoria.

Department of Transport (DOT)(2000): **NERA Report on Lorry Track and Environment Costs,** published on DOT web-site, London.

Ghaeli, Reza., Bruce, G., Hutchinson, Ralph Haas and Gillen, David. (2000): Pavement and Bridge Cost Allocation Analysis of the Ontario, Canada, Intercity Highway Network, **Transportation Research Record**, 1732, pp 99–107.

Gronau, Reuben. (1991): **Are Ghanean Roads Paying their Way?** INTUD, Working Paper 773, September.

Gronau, Reuben. (1994): Fuel Taxes and Road-User Charges in LDCs; Some Lessons from Sub-Saharan Africa, **Journal of Transport Economics and Policy,** Vol. 28, No. 3, pp. 255–73.

Hajek, Jerry J., Tighe, Susan L and Hutchinson, Bruce G. (19XX): Allocation of Pavement Damage Due to Trucks using a Marginal Cost Method, **Transportation Research Record**, 1613, pp. 50–56.

HDM 4 Manual (2001): **Analytical Framework and Model Descriptions,** Volume 4, Birmingham.

Heggie, Ian and Vickers, Piers. (1998): **Commercial Management and Financing of Roads**, World Bank Technical Paper No. 409, The World Bank, Washington, DC.

Highway Research Board (1961): **The AASHO Road Test; History and Description of Project,** Special Report 61 A, Washington D.C.

Institut für Wirtschaftspolitik und Wirtschaftsforschung (IWW), Universität Karlsruhe und Prognos AG (2002): **Wegekostenrechnung für das Bundesfernstrassennetz under Berucksichtigung der Vorbereitung einer streckenbezogen Autobahnbenutzungsgebuhr,** Auftrag des Bundesministeriums für Verkher, Bau- und Whonungswesen, Basel/Karlsruhe.

Joint Task Team (1995): **Review of the SADC/SACU Road User Charges Studies by the Joint Task Team,** Stage 1 and Stage 2 Reports, Prepared for Reference Group, SACU Road User Charges Study, Pretoria and Maputo.

Johansson, Per. and Nilsson, Jan-Eric. (2001): **An Economic Analysis of Track Maintenance,** unpublished paper available at www.du.se/~jen.

Kommunikationsdepartementet (1985): **Kostnader och avgifter inom trafiksektorn;** Rapport utarbetad av en arbetsgrupp inom kommunikationsdepartementet, Ds K 1985:2, Stockholm.

Li, Zongzhi., Kumares, C. Sinha and McCarthy, Patrick S. (2001): Methodologies to Determine Load- and Non-Load-Related Shares of Highway Pavement Rehabilitation Expenditures, **Transportation Research Record**, 1747, pp. 79–88.

Li, Zongzhi., Kumares, C. Sinha and McCarthy, Patrick S. (2002): A determination of load- and non-load shares of highway pavement routine maintenance expenditures, **Road & Transport Research**, Vol. 11, No. 2, pp. 3–12.

Lindberg, Gunnar. (2002): **Marginal cost of road maintenance for heavy goods vehicles on Swedish roads,** Deliverable 9, Case Study 5b, UNITE.

Link, Heike. (2002): **Road Econometrics – Case Study on Renewal Costs of German Motorways,** Deliverable 10, Annex A1a, UNITE.

Link, Heike., Dodgson, Johan S., Maibach, Markus and Herry, Max. (1999): **The Costs of Road Infrastructure and Congestion in Europe**, Physica Verlag, Heidelberg.

Link, Heike and Lindberg, Gunnar. (2000): **Marginal Cost Methodology for Infrastructure Costs,** Interim Report 5.3, UNITE.

Link, Heike et al. (2002): **Case Studies on Marginal Infrastructure Costs**, Deliverable 10, UNITE.

Littlechild, S.C. and Thompson, G.F. (1977): Aircraft landing fees: a game theory approach, The **Bell Journal of Economics,** Vol. 8, pp. 186–204.

Martin, Tim. (1994): **Estimating Australia's attributable road track costs,** Research Report No. 254, Australian Road and Research Board, Vermont South, Victoria.

Martin, Tim C. (2000): **Assessment of additional infrastructure costs due to increased axle mass limits**, Research Report ARR 329, ARRB, Vermont South, Victoria.

Martin, Tim. (2001): **The effect of climate on the cost of road maintenance in Australia,** Research Report ARR 353, ARRB, Vermont South, Victoria.

Ministry of Transport (1995): **National Roading Account: The Cost of Roading Infrastructure,** Land Transport Pricing Study, Wellington, N.Z.

Ministry of Works, Transport and Communication (MWTC) (1997): **A Detailed Comparison of the Namibian Road User Charges Model and the World Bank Road User Charges Model,** Windhoek, July.

National Road Traffic Commission (NRTC) (2002*):* **Updating Heavy Vehicle Charges; Draft Policy Paper,** NRTC Web-site, Melbourne.

Newbery, David M. (1985); **Road Costs Attributable to Resealing in Tunisia,** unpublished working document, Princeton University.

Newbery, David M. (1988a): Road User Charges in Britain, **The Economic Journal**, 98 (Conference 1988), pp. 161–176.

Newbery, David M. (1988b): Road Damage Externalities and Road User Charges, **Econometrica**, Vol. 56 No. 2 (March), pp. 295–316.

Newbery, David M. (1988c): Charging for Roads, **The World Bank Research Observer,** 3, no 2, July, pp. 119–138.

Newbery, David M. (1989); Cost recovery from optimally designed roads, **Economica** 56, pp. 165–85.

Newbery, David M. (1990): Pricing and congestion: Economic principles relevant to pricing roads, **Oxford Review of Economic Policy,** vol. 6, No. 2 pp. 22–38.

Newbery, David M., Gordon, A., Hughes, William, Paterson, D.O and Bennathan, Esra. (1988): **Road Transport Taxation in Developing Countries; The Design of User Charges and Taxes for Tunisia,** World Bank Discussion Papers 26, Washington DC.

Nix, Fred. (2000): **Road Infrastructure: Outlook and Policy Implications,** paper prepared for the Canada Transportation Act Review Panel, November.

OECD (1988): **Heavy Trucks, Climate and Pavement Damage,** Paris

Ozbay, Kaan., Bartin, Bekir and Berechman, Joseph. (2001): Estimation and Evaluation of Full Marginal Costs of Highway Transportation in New Jersey, **Journal of Transportation and Statistics**, April, pp. 81–103.

Ritlett, L.R., Hutchinson, B.G. and Haas, R.C.G. (1989): Cost Allocation Implications of Flexibile Pavement Deterioration Models, **Transportation Research Record**, 1215, pp. 31–42.

Rosalin, N. and Martin, T. (1999): **Analysis of Historical Data on Pavement Performance**, ARRB TR Report RE7134, Vermont South, Victoria, Australia.

Rothengatter, Werner. (2000): **Kilometer-based Charging of Heavy Goods Vehicles on German Motorways,** IWW, University of Karlsruhe, unpublished.

Sandler, Todd and Tschirhart, John T. (1980): The Economic Theory of Clubs: An Evaluative Survey, **Journal of Economic Literature,** Vol. XVIII, December, pp. 1481–1521.

SIKA (2000): **Översyn av förutsättningarna för marginalkostnadsbaserade avgifter i transportsystemet**, Slutredovisning, SIKA Rapport 2000:10, Stockholm.

SIKA (2002): **Trafikens externa effekter,** SIKA Rapport, Stockholm.

Sansom, Tony, Nash, Chris, Mackie, Peter, Shires, Jeremy and Watkiss, Paul. (2001): **Surface Transport Costs and Charges, Great Britain 1998**, Institute for Transport Studies, University of Leeds.

Small, Kenneth A., Winston, Clifford and Evans, Carol A. (1989): **Road Work, A new highway pricing and investment policy,** The Brookings Institution, Washington, DC.

Transportation Research Board (1996): **Paying our Way: Estimating Marginal Social Costs of Freight Transportation,** Special Report 246, Washington DC.

US Department of Transportation (1997): **Federal Highway Cost Allocation Study**, Washington, DC.

US Department of Transportation and Federal Highway Administration (2000): **Addendum to 1997 Federal Highway Cost Allocation Study**, Final Report, Washington, DC.

van den Bossche, Anton, et al. (2001): **Marginal Cost Methodology**, Deliverable 3, UNITE.

Vitaliano, Donald F. and Held, James. (1989): Marginal Cost Road Damage and User Charges, **Quarterly Review of Economics and Business**, pp. 32–49.

Vägverket (2000): **Marginalkostnader inom vägransportsektorn – Underlagsmaterial från Vägverket till SIKAs slutrapport gällande projektet "Översyn av förutsättningarna för marginalkostnadsprissättning inom transportsektorn,** Borlänge.

Vägverket (2001): **Översyn av marginalkostnader inom vägtransportsektorn; Underlag från Vägverket till SIKA**, December, Borlänge.

Watanatada, T. et al. (1989): **The Highway Design and Maintenance Standards Model,** Volume 1, Description of the HDM-III Model, The Johns Hopkins University Press, Baltimore.

Annex: UK Cost Allocation Method

Category of Cost	Cost allocation driver
Capital charges	
All capital charges	15% on basis of maximum gross vehicle weight , and 85% on basis of passenger car unit kms
Maintenance costs	
Long life pavements	On basis of standard axle kms
Resurfacing	On basis of standard axle kms
Overlay	On basis of standard axle kms
Surface dressing	20% on the basis of vehicle kms, and 80% on basis of average gross vehicle weight kms
Patching & minor works	20% on the basis of average gross vehicle weight kms, and 80% on basis of standard axle kms
Drainage	100% on the basis of vehicle kms
Bridges & remedial earthworks	100% on basis of average gross vehicle weight kms
Footways, cycle tracks & kerbs	50% on the basis of average vehicle weight kms, and the other 50% to pedestrians (except on motorways)
Fences & barriers	33% on the basis of vehicle kms, and 67% on the basis of gross vehicle weight kms
Grass & hedge cutting	100% on the basis of vehicle kms
Traffic signs & pedestrian crossings	100% on the basis of vehicle kms
Sweeping & cleaning	50% on the basis of vehicle kms , and the other 50% to pedestrians (except on motorways)
Road markings	10% on the basis of vehicle kms, and 90% on the basis of average gross vehicle weight kms
Winter maintenance & miscellaneous	50% on the basis of vehicle km, and the other 50% to pedestrians (except on motorways)
Policing & traffic wardens	
Police and traffic warden costs	On the basis of vehicle kms

Cost allocation drivers

Vehicle kms: unweighted vehicle kms
Passenger car unit kms: vehicle kms multiplied by PCUs
Average gross vehicle weight kms: vehicle kms multiplied by average vehicle weight
Maximum gross vehicle weight kms: vehicle kms multiplied by maximum vehicle weight
Standard axle kms: a standard axle is a measure of the relative road wear caused by different vehicles, and is equal to the sum of the fourth powers of the weights in kgs on each axle divided by 10,000. It therefore compares the road wear caused by any vehicle with the road wear caused by a ”standard axle” of ten tonnes.

Source: Department of Transport (2000), Table 5.1.