TØI report
456/1999

PB2002-104994

Econometric models of road use, accidents, and road investment decisions

Volume I

Lasse Fridstrøm

Institute of Transport Economics

Norwegian Centre for Transport Research

TØI report
456/1999

Econometric models of road use, accidents, and road investment decisions

Lasse Fridstrøm

Volume I:

Introductory overview
The barely revealed preference behind road investment priorities (Essay 1)
Measuring the contribution of randomness, exposure, weather, and daylight to the variation in road accident counts (Essay 2)

ISSN 0802-0175
ISBN 82-480-0120-2

Oslo, November 1999

Title: Econometric models of road use, accidents, and road investment decisions. Volume I

Author(s): *Lasse Fridstrøm*

TØI report 456/1999
Oslo, November 1999
88 pages
ISBN 82-480-0120-2
ISSN 0802-0175

Financed by:
Research Council of Norway (formerly NTNF), Institute of Transport Economics (TØI)

Project: 1523 Road use, accidents, and fuel consumption

Project manager: Lasse Fridstrøm

Quality manager: Marika Kolbenstvedt

Key words:
Road investments; revealed preference; cost-benefit ratio; road accidents; randomness; Poisson regression; econometrics

Summary:
This volume contains the first part of the author's Ph D dissertation, and includes an introductory overview as well as two essays. The first essay - entitled "The barely revealed preference behind road investment priorities", coauthored by Rune Elvik, and reprinted from Public Choice 92:145-168 (1997) - uses rank order logit analysis in an attempt to explain the priorities set by the regional Norwegian road administrations. The second essay - entitled "Measuring the contribution of randomness, exposure, weather, and daylight to the variation in road accident counts", coauthored by Jan Ifver, Siv Ingebrigtsen, Risto Kulmala, and Lars Krogsgård Thomsen, and reprinted from Accident Analysis and Prevention 27:1-20 (1995) - applies generalized Poisson regression analysis to pooled cross-section/time-series data covering all provinces of Denmark, Finland, Norway, and Sweden.
The third and last essay is printed in a separate Volume II (TØI report 457/1999).

Tittel: Econometric models of road use, accidents, and road investment decisions. Volume I

Forfatter(e) *Lasse Fridstrøm*

TØI rapport 456/1999
Oslo: November 1999
88 sider
ISBN 82-480-0120-2
ISSN 0802-0175

Finansieringskilde:
Norges Forskningsråd (tidl NTNF), Transportøkonomisk institutt (TØI)

Prosjekt: 1523 Vegtransport, ulykker og energi

Prosjektleder: Lasse Fridstrøm

Kvalitetsansvarlig: Marika Kolbenstvedt

Emneord:
Veginvesteringer; prioritering; nytte/kostnads-brøk; trafikkulykker; tilfeldighet; Poisson-modell; økonometri

Sammendrag:
Rapporten utgjør først bind av forfatterens dr.- polit.-avhandling og inneholder en sammenfattende innledning samt to essays. Det første essayet handler om vegkontorenes prioritering mellom investeringsprosjekter. De ulike vegprosjektenes respektive samfunnsøkonomiske lønnsomhet viser seg å ha nokså beskjeden betydning for deres innbyrdes prioritering. Det andre essayet analyserer hvordan fylkesvise ulykkestall i Danmark, Finland, Norge og Sverige avhenger av trafikkmengde, dagslys, værforhold og tilfeldigheter. En svært betydelig del av bevegelsene i dødsulykkestallene kan tilskrives rent tilfeldige variasjoner. Avhandlingens tredje og siste essay er trykt i eget bind (TØI-rapport 457/1999).

Language of report: English

The report can be ordered from:
Institute of Transport Economics, the library
P O Box 6110 Etterstad N-0602 Oslo, Norway
Telephone +47 22 57 38 00 Telefax +47 22 57 02 90
Price NOK 0

Rapporten kan bestilles fra:
Transportøkonomisk institutt, biblioteket
Postboks 6110 Etterstad, 0602 Oslo
Telefon 22 57 38 00 - Telefaks 22 57 02 90
Pris 0

Introductory overview

Contents

Preface

The present volume contains the first part of the author's dissertation for the *dr. polit.* degree at the Institute of Economics of the University of Oslo.

In total, the dissertation consists of an introductory overview and three accompanying essays.

The first essay – entitled «The barely revealed preference behind road investment priorities» and co-authored by Rune Elvik – has been published in *Public Choice* **92**: 145-168 (1997).

The second essay – entitled «Measuring the contribution of randomness, exposure, weather, and daylight to the variation in road accident counts» and co-authored by Jan Ifver, Siv Ingebrigtsen, Risto Kulmala, and Lars Krogsgård Thomsen – can be found in *Accident Analysis and Prevention* **27**: 1-20 (1995). This paper is based on the report *«Explaining the variation in road accident counts»*, by the same authors, issued by the Nordic Council of Ministers (Nord 1993:35).

Both of these essays are reprinted, with the kind permission of Kluwer Academic Publishers and Elsevier Science Ltd, respectively, in this Volume I, which also contains the introductory overview.

The third essay – entitled «An econometric model of car ownership, road use, accidents, and their severity» – is by far the largest, and printed in a separate Volume II (TØI report 457/1999).

Oslo, November 1999
INSTITUTE OF TRANSPORT ECONOMICS

Knut Østmoe
Managing Director

Marika Kolbenstvedt
Head of Department

Acknowledgements

The author is indebted to a large number of people and institutions for their help in completing the present analyses.

First and foremost, thanks are due my colleague Peter Christensen for his help in extracting seemingly endless sets of specified casualty counts from the national register of road accidents, and for his fruitful comments and advice at various stages of the research process. Without Peter's invaluable assistance from an early stage, it is doubtful if this thesis would have ever seen the light.

I was also, in an early phase of the process, greatly helped by Per Christian Svae, who set up and organized the first version of the database for essays 2 and 3, using the dBase IV format.

For our most constructive cooperation during the so-called MBAT project («Metode for Beskrivelse og Analyse av ulykkesTall»), which forms the basis for essay 2, I am indebted to my Nordic partners Jan Ifver, Risto Kulmala and Lars Krogsgård Thomsen. Thanks are also due to my Norwegian colleagues Siv Ingebrigtsen and Jan Erik Lindjord, who helped perform certain parts of the statistical and numerical calculations for the MBAT project, to Tore Vaaje, who initiated the project, and to the Nordic Council of Ministers, which financed it.

For the analysis underlying essay 1, I was helped by James Odeck and Toril Presttun of the Norwegian Public Roads Administration in obtaining a data set on road investment priorities, by Rune Elvik, Jan Erik Lindjord og Lasse Torgersen in «purging» and preparing the data for analysis, and by Inger Spangen in interpreting and qualifying the results.

This doctoral dissertation would not have been possible without the financial support – which is gratefully acknowledged – of the Royal Norwegian Council for Scientific and Industrial Research (NTNF – now merged into the Research Council of Norway), of the Norwegian Ministry of Transport and Communications, and of my employer – the Institute of Transport Economics (TØI). For the support received through TØI I am indebted to each and every employee, who have accepted to cross-subsidize my research through a large number of years and, on top of that, have offered continual encouragement and advice. Although no one is forgotten, I especially want to thank Torkel Bjørnskau, Rune Elvik, Stein Fosser, Alf Glad, Odd Larsen, Harald Minken, Arne Rideng, and Fridulv Sagberg, for their helpful hints and assistance, Astrid Ødegaard Horrisland and Anne Marie Hvaal for their superb library services, Svein Johansen and Jack van Domburg for their graphical and printing services, Lars Hansson, Jurg Jacobsen, Tore Lien, Torbjørn Strand Rødvik, and Arne Skogli for their computer support services, and Laila Aastorp Andersen, Jannicke Eble, Magel Helness, Trude Rømming, and Unni Wettergreen for their secretarial assistance.

I am generally indebted to my colleagues at the Department of Safety and Environment for our inspiring discussions and exchange of views and for their warm support in my apparently never-ending pursuit of the *dr. polit.* degree. I am particularly grateful to my Head of Department Marika Kolbenstvedt, who helped provide the resources and energy needed to give this thesis its final touch.

Certain people outside my Institute have also contributed significantly to the realization of this thesis. Morten Merg of the Road Traffic Information Council («Opplysningsrådet for vegtrafikken») provided machine-readable data on the vehicle stock by county and year. Bjørn

Reusch and the Norwegian Petroleum Institute helped provide data on fuel sales by county and month. Finn Harald Amundsen, Kjell Johansen and Lars Bendik Johansen of the Public Roads Administration were most helpful in making available a suitable data set on traffic counts. I am generally grateful to a large number of helpful researchers and employees in the Central Bureau of Statistics (now Statistics Norway), in the Norwegian Social Science and Data Services (NSD), and in the Bank of Norway, who on various occasions have provided the necessary data input and background material for my analyses. I have learnt a lot from the constructive comments and advice offered by Finn Jørgensen of the Bodø Graduate School of Business, and from my cooperation, within the COST 329 action and the OECD Scientific Expert Group RS6 on «Safety Theory, Models, and Research», with Sylvain Lassarre of the INRETS. I have also benefited greatly from the insights offered, through the AFFORD project for the European Commission, by Erik Verhoef of the Free University of Amsterdam.

Without the software support offered by Francine Dufort and the TRIO team at the Centre de Recherche sur les Transports of the Université de Montréal my analyses would not have been feasible.

Last, but not least, I wish to extend a warm thank you to my advisors Erik Biørn and Marc Gaudry, for their untiring support, encouragement and advice, and for letting me partake of their immense professional competence and patience.

Aim, motivation, and dissertation outline

Object of study

The present dissertation is concerned with road transportation in Norway, as seen in a primarily economic perspective, and examined by means of econometric methods.

Our principal focus is on the accident generating process, its determinants, and its consequences.

The accident generating process is closely linked to the mobility pattern in general and to road travel demand in particular. As a background for understanding this process, but also because the topics have considerable interest *per se*, we examine car ownership and road use demand relations, their interrelationships with road infrastructure supply, as well as the decision process determining the characteristics of this infrastructure.

A wide perspective on accidents and road use

To understand the accident generating process, it is – in our view – necessary to adopt a fairly wide perspective, paying attention to the many natural, technological, social, political, and economic background factors that have a bearing on the road accident toll.

It might be fruitful to distinguish between six broad categories of factors influencing accident counts.

First, accident numbers depend on a number of truly *autonomous factors, determined outside the (national) social system*, such as the weather, the natural resources, the state of technology, the international price of oil, the population size and structure. These are factors that can hardly be influenced (except perhaps in the very long term) by any (single) government, no matter how strong the political commitment.

Second, they depend on a number of *general socio-geographic and economic conditions*, some of which are – in practice or in principle – subject to political intervention, although rarely with the explicit purpose of promoting road safety, nor – more generally – as an intended part of transportation policy. Such factors include industrial development, (un)employment, disposable income, consumption, taxation, inflation, public education, etc.

At a third level, the size and structure of the *transportation sector*, and the policy directed towards it, obviously have a bearing on accident counts, although usually not intended as an element of road safety policy. Under the assumption of constant *risk*, the accident frequency is – by definition – proportional to the amount of *exposure*, as measured, e g, in terms of vehicle or person kilometers (i e, *road use*). Aggregate road use is, however, in our perspective not an exogenous factor; it is the result of innumerable choices made by individual private consumers, households and producers. These choices are – in turn – conditioned by certain long-term, asset ownership decisions made by private individuals, which result in a certain number of driver's licenses, a certain size and structure of the vehicle pool, and a certain spatial distribution of residence and employment. Moreover, short-term as well as long-term individual decisions are influenced by certain public policy variables, concerning, e g, public transportation level-of-service and fares, the level of fuel and vehicle taxation, and road infrastructure supply.

Fourth, accident counts are susceptible to influence – and, indeed, influenced – by *accident countermeasures*, i e measures intended to reduce the risk of being involved or injured in a road accident, as reckoned per unit of exposure.

Fifth, the accident statistics depend, of course, on the system of *data collection*. Accident underreporting is the rule rather than the exception. Changes in the reporting routines are liable to produce fictitious changes in the accident counts.

Finally, accident counts, much like the throws of a die, are strongly influenced by sheer *randomness*, producing literally unexplainable variation. This source of variation is particularly prominent in small accident counts. For larger accident counts, the law of large numbers prevails, producing an astonishing degree of long-run stability, again in striking analogy with the dice game.

Three essays

The present dissertation, which – in addition to this introductory overview – consists of three essays, is an attempt to understand and analyze road use, accidents, and road infrastructure formation in such a wide perspective.

In *essay 1* – «The barely revealed preference behind road investment priorities», co-authored by Rune Elvik – the objective is to explain the ranking of investment opportunities compiled for the four-year National Road Plan 1990-93 by the regional offices of the Public Roads Administration. This essay is fairly narrow in scope, in that it covers only a certain part of a decision process determining a limited set of conditions confined to the third level in the above list, i e to the transportation sector.

Of particular interest in this analysis is whether road investment decisions are governed by the objective to maximize total user benefit within a given budget constraint, in which case one would expect to find a highly significant coefficient for the benefit/cost ratio of an investment project. Assuming that this is the leading principle of road investment priorities, one might expect infrastructure improvements to have a comparatively strong influence on mobility and economic growth, and hence – potentially – on the accident frequency. If, on the other hand, road infrastructure improvements are primarily motivated by safety concerns, one might expect such improvements to bring the risk level down, while leaving exposure (mobility) more or less unaffected. The end result of this would be a decrease rather than an increase in the accident frequency.

In *essay 2* – «Measuring the contribution of randomness, exposure, weather, and daylight to the variation in road accident counts», co-authored by Jan Ifver, Siv Ingebrigtsen, Risto Kulmala and Lars Krogsgård Thomsen – the main purpose is to assess the importance of certain factors that are normally not controllable by policy makers. Some of these factors, such as weather and daylight, belong at the uppermost level in the list above, while exposure clearly belongs at the third level and randomness at the sixth. Many of these factors, although frequently ignored in traffic safety analyses, have potentially a very strong influence on accident counts. Our intent is to measure just how strong.

Only after all these uncontrollable factors have – so to speak – been «controlled *for*», can one interpret changes in the accident rate as attributable to the primary policy variable of interest – accident countermeasures. In other words, even if weather, daylight, randomness and exposure are not susceptible to influence by traffic safety policy measures, the identification and estimation of their effects is clearly policy relevant.

In *essay 3* – «An econometric model of car ownership, road use, accidents and their severity» – we carry the analysis of road accident counts much further, by estimating relations for aggregate car ownership, overall and heavy vehicle road use, rural and urban seat belt use, and injury accident frequency and severity. That is, rather than considering exposure as (exogenously) given, we attempt to estimate its determinants. In so doing, we take a further step back in the chain of effects, estimating the determinants of car ownership as well, since the latter variable is a predominant factor behind mode choice and travel frequency decisions.

For input into the road use and accident frequency equations, we develop an econometric method to estimate exposure (overall and heavy vehicle kilometers) from a combination of traffic counts, fuel sales statistics, fuel prices, calendar data, weather conditions, and vehicle pool characteristics.

In this essay, unlike most safety analyses, we consider dependent and independent variables belonging to all the six levels identified in the previous section.

Economics, econometrics, transportation, and road safety

Economic and econometric analyses have a long tradition within the field of transportation, but a rather short history of application within accident analysis and prevention.

Transportation is a sector characterized by frequent market failure and multiple externalities, of which accidents are but one. Examples from transportation has inspired a number of leading economists to develop theorics and methods that have later acquired a much wider area of application. Using the example of a bridge, Dupuit (1844) was probably the first economist to rigorously analyze the efficient pricing of public goods. Pigou (1920) and Knight (1924) developed their theory of externality taxation using the example of a congested road. Coase (1960) used the problem of sparkles from a railway to discuss property rights in relation to externalities.

Even certain parts of the econometric toolbox have emerged largely in response to the needs of transportation research, being brought forward by analysts with a prime interest in this field of application. Most clearly, this is the case of disaggregate discrete choice modeling, to which McFadden (1974, 1978, 1981) is generally considered to have made the single most important methodological contribution. Such modeling has become the cornerstone and leading paradigm of travel demand analysis, among planners and practitioners as well as for theorists (see Ben-Akiva and Lerman (1985) for an excellent introduction).

The use of econometric methods for the purpose of accident analysis is much less common, and by no means essential to the development of econometrics. We are, however, going to argue that although econometrics was originally developed as a toolbox for economic research, it may – in a sense – be even better suited for accident analysis.

Road accidents as an externality

It is generally acknowledged (European Commision 1996, Maddison et al 1996, Verhoef 1996, Nash 1997) that *road transportation* is an activity characterized, at least occasionally, by particularly large external costs.

Such externalities may include accidents, environmental effects, noise annoyance, congestion, and road wear. However, it should be kept in mind that these costs are generally not external in their entirety.

The issue of *road accident* externalities has been the subject of several important studies in recent years. A common theoretical finding resulting from these studies is that the external accident cost of road use is a function of the marginal relationship between road use and accidents, as expressed, for instance, by the elasticity.

However, very few studies provide well-founded empirical evidence as to the (range of) value(s) of this elasticity. In the words of Newbery (1988:171),

> «The key element in determining the accident externality cost is [...] the relationship between traffic flow and accident rates, where the evidence is sketchy, to say the least.»

An interesting question is thus whether accidents and accident victims tend to increase more or less in proportion to the traffic volume, in other words if the elasticity of casualties with respect to road use is smaller than, equal to or perhaps larger than unity. Is this elasticity perhaps not constant, but depending on the level of traffic density or «saturation»? Since, in an «unsaturated» traffic environment, the number of possible two-party conflict situations may be thought to increase in relation to the square of the number of vehicles on the road, one might imagine an elasticity much larger than one in the early phase of the automobile era. Newbery (1988) argues that, in such a case, there would be an externality involved which is at least as large as the total cost of the accident.

As roads become crowded, however, traffic density is bound to affect driving behavior, notably speed, thus forcing down the number of conflicts, or at least the severity of their outcome. Where on this curve are we? This is an empirical question that can only be resolved by means of appropriate econometric analysis, allowing for explicitly and estimably non-linear relationships.

In extending this line of reasoning, one may identify four rather intriguing questions: (i) Are we approaching the stage at which the accident externality generated by the marginal road user is zero or perhaps even positive, on account of the marginal road user's contribution to congestion and hence to speed limitation? (ii) Or are we, perhaps, in some heavily congested regions even at a stage where the *total marginal accident cost* (external *and* internal) of road use is approaching zero? (iii) Is this (one of) the reason(s) why accident counts in northwestern Europe generally have kept falling since the early 1970s, in spite of increasing road use? (iv) Is there, perhaps, some kind of trade-off between congestion and accident externalities, the sum of the two being less variable than either, since congestion tends to reduce accidents and/or their severity?

If such a «substitutability» between externalities exists, it has important implications for policy. Efforts to relieve congestion may entail not nearly the same social benefit as if these two externalities were not related, perhaps – depending on the relative values attached to time savings versus life and health – no benefit at all.

In essay 3, we estimate the partial elasticities between road use and injury accidents of varying severity. Using a pooled cross-section/time-series data set, we are able to identify separate exposure and traffic density effects and estimate their strength. Relying on a Box-Cox regression model (Box and Cox 1964), we are even able to estimate the form (curvature) of the relationship between injury accidents and exposure/density. The analysis reveals that there is probably a large accident externality generated by heavy vehicle road use, but that the marginal external accident cost of private car use is quite small, perhaps even negative.

Risk compensation in an economic perspective

Perhaps the most intriguing issue of accident and safety analysis over the last couple of decades is the question of *risk compensation*, sometimes referred to as *behavioral adaptation* or *offsetting behavior*.

In a narrow sense, risk compensation occurs when a decision maker perceives some exogenously determined *increase* in risk taking place, and changes his behavior so as to counteract, to a smaller or larger extent, this *initial* risk increase by an enhanced safety effort.

In a broader sense, one may refer to risk compensation, offsetting behavior, or behavioral adaptation, as the decision-maker's response to *any* exogenous change in risk, *positive or negative*, i e regardless of the direction of initial change. In the sequel, we shall be using the term risk compensation in this broader sense.

Few studies have aroused more controversy than the seminal paper by Peltzman (1975), who concluded that the vehicle safety design standards promulgated by the US National Highway and Traffic Safety Administration had done nothing to reduce the highway death rate. These regulations, which were imposed during the 1960's, required that new cars be equipped with (i) seat belts for all occupants, (ii) energy-absorbing steering column, (iii) penetration-resistant windshield, (iv) dual braking system, and (v) padded instrument panel.

Peltzman (1975) regressed road fatalities on a set of variables assumed to affect risky driving over the preregulatory period 1947-65, used this regression to predict traffic death rates for the postregulatory period 1966-73, and then compared the actual and predicted death rates. He found that while car occupant death rates had decreased by nearly 10 per cent, non-occupant death rates were up by some 30 per cent, leaving the overall death rates largely unaffected. Peltzman's interpretation was that drivers had reacted to the regulation by substituting «driving intensity» for safety. Although this behavioral adaptation was not large enough to completely offset the initial (engineering) effect on car occupant safety, it adversely affected pedestrians, who had not benefited from any initial safety improvement.

At about the same time, a similar but even more radical hypothesis, developed from a psychological angle, was put forward by Wilde (1972, 1975, 1982). According to his *theory of risk homeostasis*, the road user endeavors to maintain a constant (target) level of risk. A subjectively perceived *initial* increase in risk (or safety) will always induce the road user to adjust his behavior in such a way as to keep the *final* risk at the target level, i e constant.

In other words, not only does risk compensation always occur, it is also 100 per cent effective, in the sense of exactly neutralizing any extraneous changes in subjective risk. If this is true, it follows that all policy measures aimed at reducing the accident rate are bound to fail, unless they (i) attack the target level of risk, i e make the road users *want* another risk level, or (ii) are not (fully) perceived by the road users.

At the other extreme, the traditionalist (engineering) view would be that behavioral adaptation does not (by and large) take place, in other words that the total (*final*) safety effect is (approximately) equal to the engineering effect.

Is this a natural topic of research for economists? It certainly is.

Consider a utility maximizing road user whose utility function has only two arguments – speed (s) and accident risk (P):

$$U = U(s, P).$$

Assume that the marginal utilities of speed and risk are positive and negative, respectively, and that the accident risk depends on speed, as well as on some exogenous risk or safety factor x:

$$P = P(s, x)$$

The indifference map of this road user is depicted in figure 1. Utility is increasing as we move in the south-east direction. In the initial situation, the exogenous risk factor is fixed at $x = x_1$, and the road user maximizes his utility by driving at speed s_1, obtaining risk level P_1.

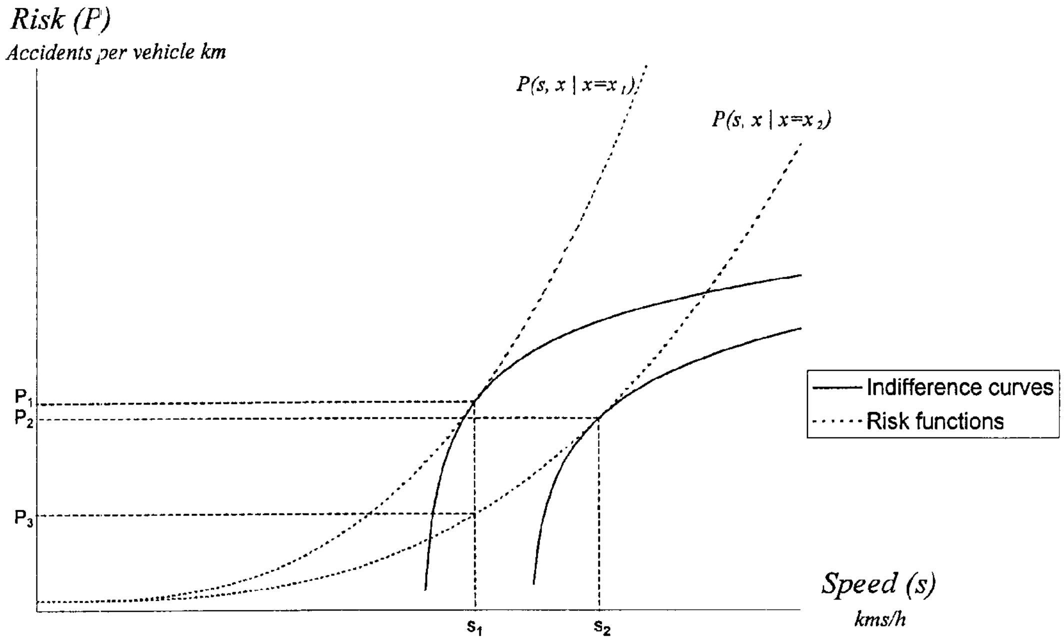

Figure 1: Indifference map of a road user

Now, suppose that the exogeneous risk factor decreases from level x_1 to x_2, shifting the $P(s, x)$ curve to the right. In this situation, speed level s_1, now resulting in risk level P_3, is no longer optimal. A much higher utility can be achieved by choosing speed level s_2, the new utility maximizing choice. Depending on the form of the indifference curves and on the function $P(s, x)$, the resulting risk level P_2 may be lower than, equal to,or higher than the initial level P_1.

This simple example serves to illustrate the more general idea, that the trade-off between risk and other utility components, such as travel time savings, excitement, or effort, is no different in essence from the trade-off between consumer goods made by a utility maximizing individual. In response to an exogenously induced change in prices, the consumer chooses a different bundle of goods, so as to still maximize his utility. It is hardly an exaggeration to say that the theory on how this *behavioral adaptation* takes place forms the very core of (neo-classical) economics.

This immediately answers the question on how an economist would typically view the issue of risk compensation. He would agree neither with (a «weak» version[1] of) the risk homeostasis hypothesis, according to which $P_2 = P_1$, nor with the traditionalist engineering view, which would imply $P_2 = P_3$.

In the economist's world, the consumer always adapts to price changes, but rarely in such a way as to keep expenditure on a given a commodity constant. This occurs only in the special case where the price elasticity of demand is exactly -1. In our diagrammatic example, the road user decides to «buy» more speed (i e, increase the speed from s_1 to s_2), as this «commodity» becomes «cheaper». However, he does not increase he consumption of speed so much as to keep the consumption of the substitute – safety – exactly constant. He decides to «buy» a little more of both[2].

By analogy, a 100 per cent risk compensation would be a mere coincidence. It is the exception rather than the rule. The same is true of zero risk compensation. The end result would typically be somewhere in-between, or perhaps – in rare cases – even outside the 0-100 per cent interval[3]. *From an economic viewpoint, the extent to which risk compensation occurs depends on the indifference map of the road user. It is therefore, in essence, an empirical question.*

In essay 3, we attempt to answer this question, in relation to a limited number of road accident countermeasures or risk factors for which relevant data have been available.

There are two types of tests by which this issue will be elucidated.

The *casualty subset test* consists in comparing effects for disjoint subsets of accidents or victims. If an *initial* safety improvement benefiting, say, car drivers is compensated, one might expect an adverse effect on other road user categories, to the extent that these are involved in bipartite or multipartite accidents with automobiles. In essence, this was the rationale behind Peltzman's (1975) controversial assertions.

A second opportunity for testing for behavioral adaptation lies in the *comparison of casualty effects by degree of severity*. While some safety (or risk) factors would tend to decrease (or increase) the *probability* of an accident, others work by influencing the *severity* of the accident, given that it takes place. One might refer to these two types of factors as *accident countermeaures* and *severity reducing measures*, respectively. Dual braking systems belong to the former group, while seat belts are an example of the latter.

A formal analysis based on utility maximization (presented in essay 3) suggests that whenever a severity reducing measure is subject to risk compensation, an increase in accident frequency

[1] In Wilde's «strong» version of the risk homeostasis hypothesis, the risk is constant *per unit of time* rather than per unit of distance traveled. We refer the reader to section 6.1.4 of essay 3 for a discussion including this case.

[2] One may view the function $P(s,x)$ as the traveler's budget constraint. The shift in this curve may be interpreted as a decrease in the «price of speed».

[3] One cannot rule out the possibility that, in another indifference map, one would have $P_2 > P_1$, or even $P_2 < P_3$. In the former case, the «price elasticity of demand for speed» is less than -1. In the latter case, which implies $s_2 < s_1$ as long as the curve $P(s,x|x=x_2)$ is fixed, speed is a Giffen good: less speed is «consumed» as the price goes down.

may be expected. And vice versa: whenever an accident countermeasure is compensated for, an increase in severity should be observed.

Thus, whenever an independent variable has an opposite sign effect on the two dependent variables (frequency and severity), there is reason to believe that we are faced with a risk or safety factor whose effect is somehow subject to compensation.

Discrete choice modeling

At the outset, discrete choice modeling was generally confined to simple, binary logit or probit analysis without much behavioral or substantive theoretical foundation (Berkson 1944, 1953, 1955). However, the theoretical developments during the 1970s paved the way for increasingly sophisticated discrete choice models with a sound and rather elegant economic basis in the form of random utility theory. The *multinomial logit* model (Theil 1969, McFadden 1974, Domencich and McFadden 1975) allows the analyst to handle several, mutually exclusive choice alternatives simultaneously. The *nested logit* model (Ben-Akiva 1973, Williams 1977) can be used to estimate choice probabilities within a hierarchy of sequential decisions, by which one is able to relax the (in)famous «Independence of Irrelevant Alternatives» (IIA) condition. The *Box-Cox logit* model relaxes the linearity assumption of the standard (linear) logit model, and allows for simultaneous estimation of logit regression coefficients and functional form (Gaudry and Wills 1978, Gaudry et al 1989).

A rather useful extension was developed by McFadden (1978), who showed that, if the number of available alternatives is intractably large, consistent estimates of the logit model parameters can be obtained on the basis of a *randomly sampled subset of alternatives*, in which the chosen alternative is included with probability one. This technique is quite useful when dealing with destination or residential choice, or – more generally – in situations where the alternatives, although discrete, are virtually innumerable, or at least rather large.

In this dissertation, we exploit the method of alternative sampling to study the *choice between competing road investment projects* within Norwegian counties (essay 1). For each four-year road plan, the public roads authorities of each county compile a full priority ranking between candidate road investment opportunities, along with cost-benefit analyses of all projects and other supplementary information. By fitting a logit decision model to this data set, we attempt to reveal the underlying preferences of the public road authorities.

A similar study, with the evocative title «The revealed preferences of a government bureaucracy», was performed by McFadden (1975, 1976) in California. Incidentally, this study also examined road investment decisions (freeway route choice), however without resorting to the alternative sampling technique or to rank-order preference data.

The demand for automobiles and road use

Transportation demand elasticities have been the subject of extensive research, at least for passenger transport (see, e g, the excellent survey articles by Oum et al. (1992) and by Goodwin (1992), and references therein). But the elasticity estimates derived are quite disparate, depending on data, functional specification, degree of aggregation, etc. Some researchers

(Goodwin 1977, Blaise 1980, Dargay 1993) suggest that consumer response may not be symmetric in regard to rising or falling prices («hysteresis»), demand being less elastic as the price (of fuel) falls than when it rises.

Few – if any – studies allow for the possibility that (aggregate) demand elasticities may not be constant over the observed range of price (or income) variation. We can, however, see no theoretical reason why they should be. Even under the (unfounded) assumption that demand elasticities with respect to the *total cost of transport* should be constant, there is every reason to think of the elasticity with respect to the *fuel price component* as variable. A higher fuel price is associated with a higher fuel cost share. If only for this reason, fuel price elasticities should be increasing (in absolute value) with the initial fuel price level. This applies to commercial freight as well as to private travel. Even in the latter case, fuel is but one of the (generalized) costs of travel incurred, other distance-dependent components being travel time, discomfort, risk, insurance, vehicle maintenance, etc.

Oum et al. (1992:153) argue cogently that

> «Different functional forms can result in widely different elasticity estimates, even with the same set of data. ... The problem is long neglected by researchers and transport practitioners. Typically, an *ad hoc* demand specification is used and little attention is directed towards testing the specification against an alternative. With the advances in econometric theory and computing technology, we think that specification testing should become an integral part of empirical transport demand research in the future.»

Being in complete agreement with this argument, we have, in essay 3, specified *estimably non-linear* demand relations, using the Box-Cox regression modeling technique. The Box-Cox class of relations is such that hyperbolic, logarithmic, linear, quadratic, cubic and higher power forms, or any power transformation in-between these, fall out as special (nested) cases within the family of generalized linear models. We will therefore be able, not only to test various specifications against each other, but also to determine the *optimal* (best fit maximum likelihood) form of the relation, as a function of the empirical evidence available.

Our suspicion is that such (Box-Cox) relations might be entirely sufficient to explain the apparent asymmetry («hysteresis») of road user response. Large price reductions tend to shift the market equilibrium into the inelastic range, while substantially increasing prices imply a movement into the highly elastic range. The theoretical and empirical insight into (the possible curvature of) these relations may have important policy implications. Gaudry and Wills (1978) have demonstrated how allowing for flexible functional forms in transportation demand relations may significantly alter the subject-matter empirical conclusions to be drawn, compared to fixed-form model specifications.

Another recommendation made by Oum et al. (1992) is this:

> «It is well known that demand becomes more elastic in the long run because users are better able to adjust to price changes. The distinction between long-run and short-run, however, is quite arbitrary in most transport demand studies. More carefully structured long-run studies are needed to integrate location choice and asset ownership decisions with transport demand.»

While localisation effects are well beyond the scope of our study, we have been able to explicitly model asset (i e, car) ownership, using a partial adjustment approach, so as to derive short term as well as long (or at least medium) term demand effects.

Most travel demand studies are based on the disaggregate discrete choice modeling paradigm. Estimates are usually based on a cross-sectional sample of individual travelers or households. Aggregate demand elasticities may be derived by means of the *sample enumeration* method, i e by simulating response at the disaggregate level and summing through the sample (Ben-Akiva and Lerman 1985). By constructing socalled *prototypical samples*, one might ensure that the disaggregate behavioral responses are weighed together in such a way as to be representative of a given population (Ramjerdi and Rand 1992).

Our approach is different. We rely on aggregate, pooled cross-section/time-series data. This approach may entail certain disadvantages in terms of aggregation and measurement errors. But a distinct advantage is that certain variables, which do not vary over a cross-section of a disaggregate units, may exhibit ample variation over time. Some of these variables, such as interest and tax rates, turn out to be quite important determinants of car ownership and use (essay 3).

Accident modeling

The Poisson probability model

There are compelling reasons to think of accidents as the outcome of a *Poisson* process (essays 2 and 3).

The first scientist to make a connection between empirical phenomena and the theoretical probability distribution derived by Poisson (1837, 1841) was Ladislaus von Bortkiewicz, who discovered that the Poisson distribution offered a perfect fit to the frequency of death by horse-kick in the Prussian army (Bortkewitsch 1898).

Bortkiewicz' observation represented an extremely original and innovative idea for his time. The relationship between probability theory and statistics, which is now seen as so obvious that teachers may have difficulty explaining the difference to their students, had not yet been recognized as a general principle applying to all probability distributions. It was, however, known that the normal distribution and the law of large numbers could be applied in this way. The elegance and usefulness of these mathematical results, associated with some of the most illustrious and prestigious mathematicians of all times (Gauss, Laplace, and others), had created a research paradigm in which almost all attention was focused on large sample theory. Against this background, the title of Bortkiewicz' book – «The law of small numbers» – was an intriguing one.

It was, however, not very accurate. We now know that the Poisson probability model is equally valid for large event counts, although the limiting distribution of the Poisson is the normal, so that in this case the distinction between the two distributions becomes immaterial (see Haight 1967, or Johnson and Kotz 1969).

Following the seminal works of Nelder and Wedderburn (1972), McCullagh and Nelder (1983), Gourieroux et al (1984a, b) and Hausman et al (1984), (generalized) Poisson regression models have come into widespread use in recent years, as applied to data sets with non-negative integer-valued dependent variables («count data»). Accident counts clearly fall into this category.

The most distinctive feature of the Poisson distribution is that its variance equals the mean. This one parameter is sufficient to uniquely determine the distribution.

A generalization of the Poisson probability model is obtained when one assumes that the Poisson parameter is itself random and drawn from a gamma distribution. The resulting compound distribution is the negative binomial (Greenwood and Yule 1920, Eggenberger and Pólya 1923).

In the negative binomial distribution, the variance generally exceeds the mean. In the limiting (special) case when it does not, we are back to the Poisson. The discrepancy between negative binomial and Poisson variance is commonly referred to as *overdispersion.*

Our essay 2 is an example of accident analysis based on generalized Poisson regression modeling. Other applications can be found in Fridstrøm and Ingebrigtsen (1991) and in Kulmala (1995), among many others.

The case for aggregate econometric accident models

Although accidents are the result of human behavioral decisions, they are not chosen. Accidents are unwanted events (except in the criminal or suicidal case). They are random and unpredictable at the micro level, in the striking sense that, had they been anticipated, they would not have happened. Each single accident is, in a sense, unpredictable by definition.

At the micro level, accidents are thus not only epistemically (subjectively) but even ontologically (objectively) random in character. Our failure to predict the single accident is not a matter of incomplete knowledge. No matter how much we learn about accident generating mechanisms or countermeasures, we would never be able to predict exactly where, when, and by whom the single accident is going to occur.

We therefore believe that accidents are governed by what Salmon (1984) has referred to as an «irreducibly statistical law», according to which single events may occur at random intervals, however with an almost constant overall frequency in the long run. Such laws are common in particle physics, but rare in the social sciences. Although the single event is all but impossible to predict, the collection of such events may very well behave in a perfectly predictable way, amenable to description by means of precise mathematical-statistical relationships. We believe that this principle applies to traffic accidents as it does to quantum physics and to the (repeated) toss of a die.

Thus, the fact that accidents are random and unpredictable at the micro level does not mean that their number is not subject to causal explanation or policy intervention. We can, through the design of road systems and vehicles and through our choice of behavior as road users, influence the *probability* of an accident occurring, thereby altering the long-term accident frequency (just as we can change the odds of the game by loading the die).

This long-term accident frequency – the *expected number of accidents* per unit of time – one might choose to think of as the result of a causal process. This process accounts for the rather striking stability observable in aggregate accident data, in which the random factors («noise», «disturbance») having a decisive effect at the micro level, are «evened out» by virtue of the law of large numbers. The causal process determines the expected number of accidents, as a function of all the factors making up the causal set (the causes).

To be specific, let ω_{tr} denote the expected number of accidents occurring during period t at location r. The expected number of accidents is, of course, not a constant – it varies with location and time, i e with r and t. One may refer to this variation, attributable to the various causal factors, as *systematic*. Unlike the random or pure chance variation, the systematic variation can – in principle – be influenced and controlled. Only the systematic variation is of interest from a policy point of view.

Let $\boldsymbol{x}_{tr} = [x_{tr1}\ x_{tr2}\ \ldots]'$ denote the vector of causal factors determining ω_{tr}, i e

$$\omega_{tr} = E[y_{tr}|\boldsymbol{x}_{tr}] = f(\boldsymbol{x}_{tr}),$$

where y_{tr} denotes the observed (factual) number of accidents at time t in location r, and $f(\boldsymbol{x}_{tr})$ is some (regression) function of the causal factors. Then, trivially,

$$y_{tr} = f(\boldsymbol{x}_{tr}) + u_{tr},$$

where the u_{tr} are disturbance terms defined as the difference between observed and expected accident counts.

In most econometric applications, the disturbance term is primarily *epistemic* – it is a reflection of the analyst's incomplete knowledge. This is, e g, the main rationale behind the random utility assumption underpinning the discrete choice theory referred to above (Ben-Akiva and Lerman 1985). There is, in general, no way to know how large the disturbance term of a given model should be expected to be, and no absolute and generally accepted yardstick against which one may judge the explanatory power (fit) of a linear logit or regression model. The goodness-of-fit measures usually computed in regression models are therefore, in our view, of very limited interest as a guide to the analyst or user.

In a perfectly specified accident regression model, however, he disturbance term may be viewed as *ontic* – a reflection of the unknowable rather than the unknown, or of the logical impossibility of casualty prediction at the micro level. Accident counts are random in a much more fundamental sense than almost any other object of study within economics or within the social sciences in general.

Therefore, the econometrician working with accident models may find himself in a very favorable position. The fact that accident counts are unpredictable at the micro level and Poisson distributed at the macro level, provides the analyst with a piece of knowledge seldom available to econometric practitioners. He knows that if he has explained all systematic variation, the remaining (random) disturbance should have variance – at each sample point – equal to the expected value of the dependent variable. In other words, having estimated the systematic part of the relation, he knows what to expect even from the random one.

The recognition of these ideas has some important implications for econometric practice.

1. The econometrician working with accident counts will have excellent knowledge of this disturbance distribution, and may apply specialized maximum likelihood estimation techniques with considerable confidence. Alternatively, if resorting to generalized least

squares or similar procedures, he will have first rate information on the heteroskedasticity of the model and hence of how to define the optimal weights[4].

2. For any accident regression model, it is possible to compute an optimal goodness-of-fit (termed P^2 in essay 2), determined by the amount of residual variation that would be left if the model had been correctly specified and all parameter estimates were equal to the true parameter values. This measure is *observable* and quite *robust* – indeed, virtually *invariant* – under alternative model specifications. By comparing the actual coefficient of determination (R^2) to P^2, the analyst is able to tell how far the model is from explaining all systematic variation. If R^2 exceeds P^2, the model is overfitted.
3. Another way to assess the performance of the model is by means of the *overdispersion parameter*, which is also calculable for any accident regression model. This parameter tends to zero as the amount of explained systematic variation approaches 100 per cent.
4. For cases in which the «pure» Poisson model is unrealistic, either because the events analyzed are not probabilistically independent, or because not all the relevant explanatory variables can be identified and measured, the *generalized* Poisson (negative binomial) maximum likelihood method represents an excellent alternative. Using this technique the analyst is able to *test*, by means of the overdispersion parameter, whether or not the pure Poisson assumption can be justified.

Thus, compared to the average econometric analyst, the accident modeler may draw upon an unusually rich and well-founded body of statistical theory. Few subjects or applications lend themselves to rigorous econometric analysis in quite the same way as road safety (essays 2 and 3).

A fifth advantage affecting the accident econometrician is the general abundance of accident statistics and the opportunity to subdivide these data into meaningful *severity classes* or *casualty subsets*, which may sometimes be used in a procedure to test for behavioral adaptation, alleged causal links, or omitted variable bias (essay 3).

A brief history of accident models

An early attempt to apply least squares regression analysis to accident rates was made by Recht (1965), who used a cross-sectional data set consisting of 45 US states as of 1960. Among the more well-known studies – not so much for its methodology as for its controversial, substantive conclusions – is Peltzman's (1975) attempt to estimate the effects of automobile safety regulation by means of aggregate time series data. Other important contributions were made by Robertson (1981), Joksch (1984), Graham and Garber (1984), Partyka (1984, 1991), Harvey and Durbin (1986), Oppe (1989, 1991a, 1991b) and Zlatoper (1984, 1987, 1989, 1991), to mention a few. Many of these studies are focused on general macroeconomic variables and their relation to accident rates.

The partial relationship between traffic volumes and accident counts has been addressed by several researchers, although rarely with sophisticated econometric techniques. An influential

[4] An iterative or at least two-step procedure may be necessary to achieve this.

set of studies were made by Smeed (1949, 1955, 1974), who generally claimed the fatality count to be proportional to the cubic root of the motor vehicle stock multiplied by the squared population. Subsequent studies have provided less clear-cut results[5].

A most important step forward was made with the DRAG[6] model for Quebec (Gaudry 1984), whose novelty consisted, *inter alia*, in (i) a substantially extended set of explanatory factors, (ii) a multi equation modeling approach, in which not only the accident frequency, but also their severity and the underlying amount of exposure were treated as endogenous variables to be explained, and (iii) an estimation technique allowing for estimably flexible (non-linear) functional forms for several dependent and independent variables.

Later modeling efforts within the DRAG tradition include a German model (Gaudry and Blum 1993), a French model (Jaeger and Lassarre 1997), a local Swedish model (Tegnér and Loncar-Lucassi 1996), an updated model for Quebec (Gaudry et al 1995), a model for California (McCarthy 1999), and the Norwegian model (TRULS[7]) to be presented in this dissertation (essay 3). An account of all of these models is forthcoming in Gaudry and Lassarre (1999).

---

[5] An excellent review of early works can be found in Satterthwaite (1981).

[6] DRAG is an acronym for «Demande Routière, Accidents et leur Gravité», or «Demand for Road use, Accidents, and their Gravity» in its English version.

[7] TRULS is an acronym for «TRafikk, ULykker og Skadegrad», meaning «Traffic, Accidents, and Severity».

Essay 1: The barely revealed preference behind road investment priorities

Background and objective

The decision process governing the allocation of resources between competing road investment projects has been the subject of several recent studies in Scandinavia (Elvik 1993, Odeck 1991 and 1996, Jansson and Nilsson 1989, Nilsson 1991, Nyborg and Spangen 1996). In general, these studies have shown very weak – if any – association between the priority ranking assigned to a given road investment project and the project's calculated economic cost, benefit, or benefit/cost ratio. This may seem surprising in view of the fact that, if a maximum economic benefit is to be obtained within the constraint of a given investment budget, an optimal decision rule would be to rank the projectsaccording to a decreasing benefit/cost ratio, and then carry out the projects in that order, until the budget is depleted.

In essay 1, we set out to reexamine this issue by means of potentially more powerful statistical methods than have previously been adopted. When previous studies have been unable to detect any clear association between benefit/cost ratio and priority ranking, could the reason be that these studies fail to take into account («control for») certain fairly important constraints to which decision makers are subject? Would a different picture emerge through an appropriate, multivariate method of analysis, in which one estimates the *partial* effect of economic cost and benefit *conditional* on the relevant constraints? Would it be possible to separate out the effects of different benefit *components*, estimating, e g, the weight put on safety improvements as compared to time savings? Are there a lot qualitative factors at play, beyond those taken account of in the benefit/cost calculations, which influence decision-making? If so, would it be possible to identify these factors?

Data and method

For the National Road Plan 1990-93 (as for any four-year road plan), each regional office of the Norwegian Public Roads Administration was required to conduct cost-benefit analyses of all candidate investment projects and to make a formal (rank order) list of priorities among these projects. From this, the central agency of the Public Roads Administration assembled an investment project data base containing all projects ranked by the regional road agencies (some 700 projects). A fairly large number of variables characterizing each project were recorded in the data base, including cost, benefit (four components), type of road, type of area, legal planning requirements, etc. By courtesy of the Public Roads Administration, we were granted access to this data base.

Given that all n_r (say) projects within a given county r have been ranked with respect to each other, a full information method of discrete choice analysis is to consider every project except the last one as «preferred» to all lower ranked alternatives. This method is known as «*exploded logit*» analysis: one «explodes» a ranking of n_r alternatives into $n_r - 1$ implicit choices, each alternative being considered «chosen» from the set including itself and all lower ranked options.

In our case, a full-fledged, exploded logit modeling procedure would require the estimation of discrete choice models with up to 99 alternatives (since, in one county, there are no less than 99 projects ranked). Given the constraints of the computer software available, such a large number of alternatives would force us to severely limit the number of independent variables used. We therefore decided to combine the exploded logit approach with another specialized discrete choice modeling technique – that of *alternative sampling*.

For each alternative, we drew a random sample of size 10 from the set of lower ranked options, thus forming – together with the «chosen» (highest ranking) project – an 11-alternative choice set. Obviously, when there were less than 10 projects left on the list, the choice set became smaller, consisting of only two alternatives for the second last option ranked (project $n_r - 1$).

A logit model with exclusively generic coefficients was then estimated on the basis of a pooled data set for all counties, each county being represented by $n_r - 1$ choice observations from a set of (up to) 11 alternatives.

Based on this model, we are – in line with common practice within, e g, value-of-time studies – able to derive implicit, marginal rates of substitution between the various pecuniary or non-pecuniary independent variables, as revealed by the choices made by the public decision maker.

Main results

Relying on our relatively information efficient method of analysis, we do find – unlike previous studies – statistically significant effects of cost and benefit on priority ranking. Both effects have the expected sign.

The effects are, however, not very large. Compared to certain other independent variables entering the model, the impacts of cost or benefit appear to be rather marginal.

Cost savings appear to be valued at about twice the rate of benefits. Even large increases in benefit appear to have an only minor effect on the probability of being ranked above a competing project. Apparently, decision makers are more concerned with geographic distribution (equity) than with allocative efficiency.

When benefit is decomposed into its various parts, *road user* benefits (primarily time savings) are seen to be by far the most important component, in terms of its size (compared to the other components) as well as by with its marginal impact on priority ranking. *Safety* benefits have a statistically insignificant effect, along with *noise abatement* benefits. Benefits accruing to the *road owner* (i e, the Government, as represented by the Public Roads Administration itself) come out with a counterintuitive, barely significant, negative effect on ranking. These benefits may take the form, e g, of reduced maintenance expenditure.

In total, the model explains no more than a small share of the variation present in the data set. The explanatory power of cost and benefit appears particularly modest. We are, in other words, «barely» able to reveal the preferences of the decision makers.

Essay 2: Measuring the contribution of randomness, exposure, weather, and daylight to the variation in road accident counts

Background and objective

Road accident statistics are, understandably, the subject of considerable interest on the part of media, policy-makers, organizations, and the general public. Instances in which accident counts are, for some reason, unusually high, receive particular attention. Such cases are almost invariably interpreted as a change in the underlying accident risk, and tend to generate some form of public action or outcry.

But accidents counts are, as already noted, influenced by numerous factors other than the risk level. The aim of this research, which was commissioned by the Nordic Council of Ministers as an inter-Nordic endeavor, was to assess how much variation in the accident counts is typically attributable to *randomness*, to *exposure*, to *weather and daylight*, and to (changes in the) accident *reporting routines*. Only when all of these factors have been controlled for can we interpret changes in accident counts as attributable to changes in risk, i e in the expected (long-term) number of accidents or victims per unit of exposure.

Method

To analyze these issues, combined cross-section/time-series data bases were established for each of the four greater Nordic countries. Monthly accident counts were recorded for each county (province), of which there are 14 in Denmark, 11 in Finland, 19 in Norway, and 24 in Sweden. The time period of observation used for this study covered between 132 (Denmark) and 168 (Norway) months, yielding at least 1 700 units of observation for each of the four countries.

Apart from accident statistics, the data bases include, *inter alia*, data on fuel sales (a proxy for exposure or traffic volume), weather conditions, the duration of daylight, changes in legislation and reporting routines (dummies), a trend variable, and dummy variables for the different counties and months.

Due to dissimilarities with respect to the availability and quality of statistical sources in the four countries, it has not been possible to adopt exactly the same variable definitions and classifications in all countries, nor has it been possible to lump all data into one, four-country data base. Thus, we have not able to analyze the variation *between* countries; only the temporal and spatial variation *within* each country has been subject to study.

Generalized Poisson (i e, negative binomial) regression models for each country were estimated with three types of dependent variables: the number of *injury accidents*, the number of *fatal accidents*, and the number of *fatalities* (road users killed). The first category includes even the fatal accidents.

Exploiting the Poisson assumption (i e, the equality between mean and variance), we show how it is possible to define goodness-of-fit measures that take account of the objective randomness inherent in accident counts. These measures may be said to compare empirical accident models, not with a bound implying that all variation is explainable, but with a yardstick implying that all *epistemic* (subjective) errors have been removed, while the *ontic* (objective) randomness remains, as an unavoidable feature of accident counts.

Five different goodness-of-fit measures of this kind were defined and calculated. One is based on the familiar coefficient of variation (R_P^2), a second (R_{PW}^2) on its weighted analogue, a third (R_{PFT}^2) on the so-called Freeman-Tukey deviates (Freeman and Tukey 1950), a fourth (R_{PE}^2, the «Elvik index») on the overdispersion parameter, and a fifth (R_{PD}^2) on the log-likelihood ratio. By relying on more than one measure we attempt to minimize the risk of drawing conclusions on account of methodological choices rather than subject matter relationships.

Main results

The five different ways of measuring explanatory power were seen to yield reassuringly similar results, with one possible exception. The measure R_{PW}^2, which is based on weighted (variance stabilizing) residuals and correspond to the Pearson chi-square statistic, is not invariant under alternative assumptions regarding the «true» («benchmark») probability model, and sensitive to estimation errors affecting the smallest accident counts. Its use is therefore discouraged.

The comparison between injury accident and fatal accident models reveals that the scope for normal random variation is strongly dependent on the size of the unit of observation, as measured by the expected number of events. For data sets in which the expected number of events is small – say, always less than 10 – a major part of the variation will typically be due to sheer chance. It is useful for the analyst to be aware of the fact that, in such cases, no model should attempt or be able to explain more than a smaller part of the observed variation. When the effects of policy interventions are to be evaluated, it is essential to be able to control for the sometimes very important random component in casualty counts.

Thus, in the models for fatal accidents per county and month, randomness accounts for between 50 per cent (Denmark) and 80 per cent (Norway) of the total variance. The difference between Denmark and Norway is simply due to the fact that the Danish counties are generally larger (in terms of population, exposure, or accidents).

For injury accidents, which are a lot more frequent, randomness accounts for less than 10 pent of the variance.

When the purely (ontologically) random variation has been subtracted, exposure (literally, gasoline sales) accounts for more than 70 per cent of the remaining (systematic) variation in injury accident counts, and more than 50 per cent of the variation in fatal accident counts. Changes in reporting routines explain up to 7 per cent, while weather and daylight may account for another 6 per cent. Taken together, our four general factors (randomness, exposure, reporting, and weather/daylight) typically explain around 90 per cent of the total (random and systematic) variation across counties and months in a Nordic country.

For casualty counts that are not probabilistically independent, such as fatalities or victims, of which there may be several in one accident, our specialized goodness-of-fit measures are somewhat harder to interpret, and generally lower. This is so because when there is (probabilistic) dependence between disaggregate events, overdispersion must always be expected[8].

[8] When the correlation between the events is negative, *underdispersion* – i e a variance lower than the Poisson mean – may occur. *Parity*, i e the number of children born to each woman, is an empirical example of this. The more children have already been born, the smaller is the probability of another birth (Winkelmann and Zimmermann 1992).

Essay 3: An econometric model of car ownership, road use, accidents, and their severity

Overview of the model TRULS

Essay 3 is by far the largest and most important of the three essays. Here, we report – in considerable detail – on our endeavor to develop a large econometric model for car ownership, road use, accidents, and their severity. The model has received the acronym TRULS[7].

The TRULS model is a member of a larger family of models, all inspired by the DRAG model for Quebec. The common features of all members of the DRAG family is an at least *three-layer recursive structure of explanation*, involving road use, accident frequency, and severity, and an econometric technique – called *BC-GAUHESEQ* (*B*ox-*C*ox *G*eneralized *AU*toregressive *HE*eteroskesdastic *S*ingle *EQ*uation) – allowing for estimably non-linear relationships (Gaudry et al 1993, Liem et al 1993).

Road use (traffic volume) is not considered an exogeneous factors, but explained by a number of socio-economic, physical and political variables. *Accident frequency* is modeled depending on road use, the presumably single most important causal factor. *Accident severity* is modeled as the number of severe injuries or fatalities per accident, i e as the conditional probability of sustaining severe injury given that an accident takes place.

Thus, the total number of fatalities (e g) is decomposable into two parts: the number of accidents × the number of fatalities per accident. This multiplicative decomposition allows for added insights and interesting substantive interpretations, as we shall see later on.

Some DRAG-type models include additional layers of explanation or prediction. The TRULS model, e g, includes (i) *car ownership*, (ii) *seat belt use*, and (iii) a *decomposition between light and heavy vehicle road use*, adding to the set of econometric equations.

Also, while most DRAG-type models use the fuel sales as a (rather imperfect) measure of the traffic volume, in TRULS we have constructed (iv) a *submodel designed to «purge» the fuel sales figures of most nuisance factors* affecting the number of vehicle kilometers driven per unit of fuel sold. These nuisance factors include vehicle fuel economy, aggregate area-wide vehicle mix, weather conditions, and fuel hoarding due to certain calendar events or price fluctuations.

A further point at which the TRULS model differs from other members of the DRAG family, is by the estimation of (v) *separate equations for various subsets of casualties* (car occupants, seat belt non-users, pedestrians, heavy vehicle crashes, etc). These equations are meant to shed further light on the causal mechanisms governing accidents and severity. In order to avoid, to the largest possible degree, spurious correlation and omitted variable biases, we develop certain *specificity tests* not previously used within the DRAG modeling framework. We refer to these tests as *casualty subset tests*.

Unlike other DRAG family models, the TRULS model starts from an assumption that casualty counts in general follow a (generalized) Poisson distribution (see Fridstrøm et al 1993, 1995). To enhance efficiency, in the accident equations we therefore rely (vi) on a *disturbance variance specification approximately consistent with the Poisson law*. To this end, we develop a special statistical procedure, termed *Iterative Reweighted POisson-SKedastic Maximum Likelihood* (*IRPOSKML*), for use within the general BC-GAUHESEQ statistical framework.

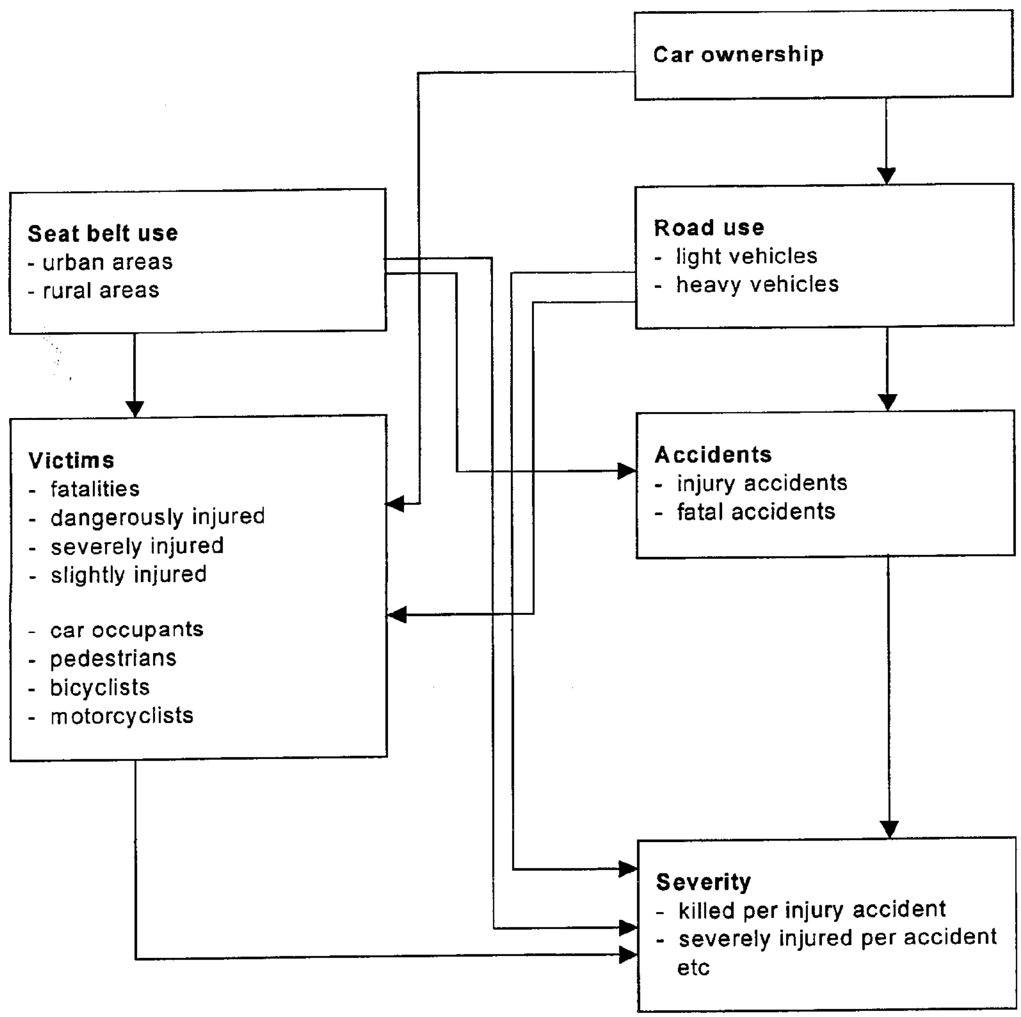

Figure 2: Dependent variables in the model TRULS

Finally, the TRULS model is the only DRAG-type model so far being based (vii) on *pooled cross-section/time-series data*. Other DRAG family models rely exclusively on time-series. Our data, however, are monthly observations pertaining to all counties (provinces) of Norway. The period of observation extends from January 1973 until December 1994, thus covering 264 months. Since there are 19 counties in the country, the data set contains a total number of 5 016 units of observation.

The structure and interdependencies between *endogenous* (*dependent*) variables in the TRULS model are shown in figure 2. In table 1 we provide an overview of (broad categories of) *independent* variables entering the model.

Note that only *direct* effects are ticked off in this table. In general, the total effect of an independent variable on – say – accident frequency, will be a sum of direct and indirect effects, as channeled though the recursive system pictured in figure 2. For instance, the interest level has a direct effect on car ownership only. However, since car ownership affects road use, which in turn affects accidents, interest rates may turn out as an important *indirect* determinant of road casualties. The tracing of such effects is the very purpose of our multi-layer modeling approach.

Table 1: Independent variables in the model TRULS

Independent variable	Direct effect upon (dependent variable)					
	Car ownership	Vehicle kms	Seat belt use	Accidents	Victims	Severity
Exposure				√	√	√
Infrastructure	√	√		√	√	√
Road maintenance				√	√	√
Public transportation	√	√		√	√	√
Population	√	√		√	√	√
Income	√	√				
Prices	√	√				
Interest rates	√					
Taxes	√	√				
Vehicle characteristics		√	√	√	√	√
Daylight		√		√	√	√
Weather conditions		√		√	√	√
Calendar effects		√		√	√	√
Geographic characteristics	√	√	√	√	√	√
Legislation			√	√	√	√
Fines and penalties			√			
Access to alcohol				√	√	√
Information		√	√			
Reporting routines				√	√	√
Randomness and measurement errors	√	√	√	√	√	√

An econometric method to estimate vehicle kilometers from traffic counts

The number of motor vehicle kilometers driven per county and month is probably the most important systematic factor behind the accident counts (cf essay 2). To arrive at a maximally accurate and reliable measure of this crucial factor, an econometric submodel was developed (chapter 3 of essay 3).

A variety of statistical indicators exist on road use, none of which, however, provide complete and accurate information. *Traffic counts* may be taken regularly at given cross-section points of the road network, but their representativity as applied to a given geographic area (say, a province or country) is, at best, hard to establish. *Fuel sales* statistics are influenced by weather variations, speed or congestion levels, hoarding, and interprovincial travel, as well as by changes in vehicle fuel efficiency.

However, by integrating many of these data sources within a consistent and fairly general econometric framework, we derive apparently reliable estimates on the number of *overall* and *heavy vehicle kilometers driven* by county and month in Norway during 1988-94. Values are then extrapolated back to 1973, exhibiting a quite acceptable degree of precision as evaluated against nation-wide annual statistics. Fuel consumption per vehicle kilometer is shown to be strongly dependent on the air temperature, rising more and more sharply as the temperature falls.

A model of aggregate car ownership and road use

Chapter 4 of essay 3 deals with aggregate car ownership and road use demand equations. These two – car ownership and use – are closely connected variables, in the trivial sense that one cannot (usually) use a private car unless the household owns one. At the same time, very few households would acquire a(n extra) car unless they intend to use it.

From a more formal, microeconomic perspective, it may be argued that car ownership is connected with certain fixed costs. These fixed cost of car ownership will be worthwhile to the individual or household only if the utility derived from the (sub)optimal[9] number of kilometers driven exceeds the variable cost. Thus, the variable cost of car *use* affects car *ownership*. And – vice versa – the fixed cost of car *ownership* affects the demand for car *use*. Elegant, discrete-continuous microeconometric models of these joint decisions have been developed by Train (1986), de Jong (1990) and Ramjerdi and Rand (1992).

Unfortunately, the mathematical structure of these microeconomic models is such as to break down should the income elasticity of demand for cars approach unity or – indeed – be even larger. Also, it might be argued against these models that the close connection between car ownership and use follows with necessity from the mathematical model structure, rather than being testable by means of empirical data.

Our approach, which is based on aggregate cross-section/time-series data, does not have these shortcomings. Also, the time-series dimension of our data set allows us to estimate the effects of certain variables that usually do not vary across a cross-section of household, which is the

[9] That is, optimal given the car ownership decision.

usual source of data for microeconometric modeling. Such variables include interest and tax rates.

A vehicle pool is an inert matter, comparable to a human population, although with generally higher rates of turnover and shorter life expectancy. The *stock* of cars registered within a given geographic unit changes from one year to the next, in response to the *flows* of (i) new car acquisition («births»), (ii) used car sales («migration»), and (iii) scrapping («deaths»). Given the very high level of purchase tax imposed on automobiles in Norway, used cars can be sold abroad only at very substantial losses. Thus, the only important downward adjustment mechanism operating at the macro level is scrapping, something which also involves heavy losses unless the car is old enough to have lost most of its market value. Hence, in the aggregate, car owners can be expected to adjust only slowly to changes in economic variables.

We therefore model car ownership as a partial adjustment process, implying that the aggregate car stock, when subject to exogenous shocks, adjusts only slowly towards its new long-term equilibrium.

Next, the aggregate demand for road use is estimated using the size of the private car pool as one important explanatory variable. The elasticity (of vehicle kilometers with respect to aggregate car ownership) turns out to be close to one.

Based on these estimates, short and long term road use demand and Engel curves are derived, the long term effects incorporating – by definition – changes affecting car ownership equilibrium. In the long term, price and income effects are apparently much stronger than in the short term. Also, demand relations are revealed – thanks to the Box-Cox modeling approach – to be strongly non-linear, yielding, e g, much more elastic demand in the higher fuel price ranges than in the lower. Explanatory variables used include road infrastructure, public transportation supply, population, income, fuel prices, vehicle prices, interest level, weather and climate, calendar effects, and geographic characteristics.

The long term income elasticity of demand for road use appears to be somewhat larger than unity, and apparently increasing with the income level. This finding may seem to have important and rather discouraging implications with respect to the goal of sustainable mobility. There is no sign that the growth in aggregate road use demand will be tapering off as the economy continues to grow – rather the contrary.

An auxiliary model of seat belt use

Seat belts are perhaps the single most important road safety measure introduced in industrialized societies in the postwar period. Studies suggest that seat belts may cut the injury or death risk by some 50 per cent, perhaps even more (Elvik et al 1989, 1997).

Seat belt wearing is to a large extent conditioned by exogenous, politically determined laws and regulations. Largely on account of these measures, seat belt use in Norway has risen from around 20-30 per cent in the early 1970s to about 80 per cent in the 1990s, according to roadside surveys. This large increase may be thought to have had a considerable effect on the road casualty toll, explaining, perhaps, in large part the downward casualty trend observed in the 1970s and -80s.

Seat belt use is not generally an observed variable (at the level of county and month during 1973-1994), although a fairly large number of roadside sample surveys exist, splitting the car drivers passing a given point on a certain day between seat belt users and non-users. However, these estimates are subject to random sampling error and also, as applied to our countywide, monthly units of observations, to an incalculable systematic error originating from the non-random sampling of roadside survey points. Fortunately, these survey points remain fixed from one survey to the other, so that the temporal *variation* in estimated seat belt use frequency is not, in the same way as its *level*, affected by the process of survey point determination.

By fitting a model to this incomplete set of observations and then imputing values for *all* units of observations in our cross-section/time-series data set, we obtain a fairly well-founded set of measures on seat belt use by county and month, in which sampling errors have been «smoothed out» and the structural information on exogenous laws and regulations has been exploited and incorporated. This submodel is presented in chapter 5 of essay 3.

A formal microeconomic analysis of risk compensation

In chapter 6 of essay 3, we develop the risk compensation argument sketched above into a more comprehensive, mathematical theory of road user behavioral adaptation. Our derivation builds on the formulation put forward by Blomquist (1986), but attempts to carry this analysis several steps further

(i) by formalizing and examining certain hypotheses put forward by Bjørnskau (1994), according to which behavioral adaptation depends on the initial risk level and expected loss, so that, e g, severity reducing measures will be compensated only if the conditionally expected loss in the event of an accident is small,

(ii) by deriving certain empirically testable implications, such as the hypothesis that if risk compensation occurs, one might expect opposite sign effects on accident frequency and severity, respectively, and

(iii) by examining not only the risk neutrality case, but also – to some extent – the risk aversion case, as defined in terms of a two-moment utility function.

Unfortunately, our data set is not sufficiently complete for full-fledged empirical tests of the many hypotheses emanating from the formal mathematical analysis. Most importantly, data are lacking on material damage accidents and on key behavioral instruments, such as speed. A number of interesting opportunities for future research are, however, indicated.

An empirical analysis of casualty counts

The core of essay 3 is a comprehensive econometric analysis of casualty counts in all Norwegian counties. Separate equations are estimated for

- injury accidents,

- car occupants injured,
- motorcycle occupants injured,
- bicyclists injured,
- pedestrians injured,
- severity of degree two[10] (severely injured per injury accident),
- severity of degree three (dangerously injured per injury accident), and
- mortality («severity of degree four», i e death victims per injury accident).

By definition, the number of fatalities is calculable as the number of injury accidents times the mortality measure. Similar multiplicative decompositions apply to the number of severely or dangerously injured, respectively. By specifying a model based on such decompositions, certain important insights may be gained, especially as regards the question of behavioral adaptation.

Independent variables in the casualty count equations include exposure (overall and heavy vehicle traffic volume, motorcycle use, public transportation supply), road infrastructure (length of network, capital invested, and maintenance), population characteristics, vehicle pool characteristics, weather and daylight, calendar effects, legislative measures, seat belt use, access to alcohol, and reporting routines. Approximately 50 independent variables are used in each equation.

Accident numbers appear to be roughly proportional to the number of vehicle kilometers, under the assumption that the road network is extended at a rate corresponding to the traffic growth. When, more realistically, the road network remains basically unchanged, casualties appear to grow at a lesser rate than the traffic volume, implying decreasing risk per vehicle kilometer, presumably because increased congestion forces speed levels down.

The analysis illustrates that it may be fruitful to see exposure as a multidimensional variable, decomposing road use by type of vehicle, and acknowledging the importance of pedestrian and bicyclist exposure, and of the public transportation system generating such exposure as a result of access and egress trips. All of these exposure components have an effect on the injury accident frequency.

Deriving compound elasticities within the recursive TRULS structure

In the last chapter of essay 3, we exploit the recursive structure of the TRULS equations to derive a complete set of compound elasticities, showing how the various casualty counts depend on any one of the exogenous variables entering the car ownership, road use, seat belt use, accident frequency, or severity equations. Direct and indirect effects are accumulated and

[10] Severity of «degree one», the number of injuries per (injury) accident, exhibits very little variation, as the number of injuries is strongly correlated with the number of injury accidents. Due to the lack of information on material damage accidents, we cannot, therefore, perform meaningful analysis of this severity level.

expressed in terms of elasticities as evaluated at the subsample means over the last year of observation (1994). A formal mathematical framework for the calculation of compound elasticities is presented.

As a result of this, we are able to assess the ultimate road casualty effect, not only of variables entering the casualty equations directly, but also of those variables exerting influence indirectly, through their impact on car ownership, road use, or seat belt use. Important examples of such variables are fuel and vehicle prices, interest rates, and tax rates.

Literature

Ben-Akiva M (1973): Structure of passenger travel demand models. PhD dissertation, Department of Civil Engineering, MIT, Cambridge, Mass.

Ben-Akiva M and Lerman S R (1985): *Discrete choice analysis: theory and application to travel demand.* MIT Press, Cambridge, Mass.

Berkson J (1944): Application of the logistic function to bioassay. *Journal of the American Statistical Association* **39**:357-65

Berkson J (1953): A statistically precise and relatively simple method of estimating the bioassay with quantal response, based on the logistic function. *Journal of the American Statistical Association* **48**:565-599

Berkson J (1955): Maximum likelihood and minimum X^2 estimates of the logistic function. *Journal of the American Statistical Association* **50**: 130-162

Bjørnskau T (1994): Hypoteser om atferdstilpasning (risikokompensasjon). Working paper TST/0512/1994, Institute of Transport Economics, Oslo.

Blaise J H (1980): Hysteresis in travel demand. *Transportation Planning and Technology* **6**(2).

Blomquist G (1986): A utility maximizing model of driver traffic safety behavior. *Accident Analysis & Prevention* **18**(5):371-375

Bortkewitsch L von (1898): *Das Gesetz der kleinen Zahlen.* B G Teubner, Leipzig.

Box G E P & Cox D R (1964): An analysis of transformations. *Journal of the Royal Statistical Sociey B* **26**:211-243.

Coase R H (1960): The problem of social cost. *Journal of Law and Economics* **3**:1-44.

Dargay J M (1993): Demand elasticities: a comment. *Journal of Transport Economics and Policy* **27**:87-90.

Domencich T & McFadden D (1975): *Urban travel demand – a behavioral analysis.* North-Holland, Amsterdam.

Dupuit J (1844): On the measurement of the utility of public works. In: Murphy D (ed): *Transport.* Penguin, London.

Eggenberger F & Pólya G (1923): Über die Statistik verketteter Vorgänge. *Zeitschrift für angewandte Mathematik und Mechanik* **1**:279-289.

Elvik R (1993): *Hvor rasjonell er trafikksikkerhetspolitikken?* Report 175, Institute of Transport Economics, Oslo.

Elvik R, Mysen A B & Vaa T (1997): *Trafikksikkerhetshåndbok: oversikt over virkninger. kostnader og offentlige ansvarsforhold for 124 trafikksikkerhetstiltak.* Institute of Transport Economics, Oslo.

Elvik R, Vaa T & Østvik E (1989): *Trafikksikkerhetshåndbok: oversikt over virkninger. kostnader og offentlige ansvarsforhold for 84 trafikksikkerhetstiltak.* Institute of Transport Economics, Oslo.

European Commision (1996): Towards fair and efficient pricing in transport. *Bulletin of the European Union,* Supplement 2/96.

Freeman M F & Tukey J W (1950): Transformations related to the angular and the square root. *Annals of Mathematical Statistics* **21**:607-611.

Fridstrøm L, Ifver J, Ingebrigtsen S, Kulmala R & Thomsen L K (1993): *Explaining the variation in road accident counts.* Report **Nord 1993:35**, Nordic Council of Ministers, Copenhagen/Oslo.

Fridstrøm L, Ifver J, Ingebrigtsen S, Kulmala R & Thomsen L K (1995): Measuring the contribution of randomness, exposure, weather and daylight to the variation in road accident counts. *Accident Analysis & Prevention* **27**(1):1-20.

Fridstrøm L & Ingebrigtsen S (1991): An aggregate accident model based on pooled, regional time-series data. *Accident Analysis & Prevention* **23**(5):363-378.

Gaudry M (1984): DRAG, un modèle de la Demande Routière, des Accidents et de leur Gravité, appliqué au Québec de 1956 à 1982. Publication 359, Centre de Recherche sur les Transports (CRT), Université de Montréal

Gaudry M & Blum U (1993): Une présentation brève du modèle SNUS-1 (Straßenverkehrs-Nachfrage, Unfälle und ihre Schwere). *Modélisation de l'insécurité routière.* Collection Transport et Communication no 47:37-44, Paradigme, Caen.

Gaudry M, Duclos L-P, Dufort F & Liem T (1993): TRIO Reference Manual, Version 1.0. Publication 903, Centre de Recherche sur les Transports (CRT), Université de Montréal

Gaudry M, Fournier F & Simard R (1995): DRAG-2, un modèle économétrique appliqué au kilometrage, aux accidents et à leur gravité au Québec: Synthèse des résultats. Société de l'assurance automobile du Québec

Gaudry M, Jara-Díaz S R & Ortúzar J de D (1989): Urban travel demand: the impact of Box-Cox transformations with nonspherical residual errors. *Transportation Research B* **23**:151-158.

Gaudry M & Lassarre S (eds) (1999): *Structural Road Accident Models: The International DRAG Family.* Elsevier (forthcoming)

Gaudry M & Wills M I (1977): Estimating the functional form of travel demand models. *Transportation Research* **12**(4):257-289.

Goodwin P B (1977): Habit and hysteresis in mode choice. *Urban Studies* **14**:95-98.

Goodwin P B (1992): A review of new demand elasticities with special reference to short and long run effects of price changes. *Journal of Transport Economics and Policy* **26**:155-169.

Gourieroux C, Monfort A & Trognon A (1984a): Pseudo maximum likelihood methods: theory. *Econometrica* **52**(3):681-700.

Gourieroux C, Monfort A & Trognon A (1984b): Pseudo maximum likelihood methods: application to Poisson models. *Econometrica* **52**(3):701-720.

Graham J D & Garber S (1984): Evaluating the effects of automobile safety regulation. *Journal of Policy Analysis and Management* **3**:206-224.

Greenwood M & Yule G U (1920): An enquiry into the nature of frequency distributions to multiple happenings, with particular reference to the occurrence of multiple attacks of disease or repeated accidents. *Journal of the Royal Statistical Society A* **83**:255-279.

Haight F A (1967): *Handbook of the Poisson distribution.* Wiley, New York.

Harvey A C and Durbin J (1986). The effects of seat belt legislation on British road casualties: a case study in structural time series modelling. *Journal of the Royal Statistical Society A* **149**:187-227.

Hausman J, Hall B H & Griliches Z (1984): Econometric models for count data with an application to the patents-R&D relationship. *Econometrica* **52**(4):909-938.

Jaeger L & Lassarre S (1997): Pour une modélisation de l'évolution de l'insécurité routière. Estimation du kilométrage mensuel en France de 1957 à 1993: méthodologie et résultats. Rapport DERA no 9709, Convention DRAST/INRETS, Strasbourg/Paris.

Jansson J O & Nilsson J-E (1989): Spelar samhällsekonomiske kalkyler någon verklig roll i vägväsendet? *Ekonomisk Debatt* no 2:8595.

Johnson N L & Kotz S (1969): *Discrete distributions*. Wiley, New York.

Joksch H C (1984): The relation between motor vehicle accident deaths and economic activity. *Accident Analysis & Prevention* **16**:207-210.

Jong G de (1990): An indirect utility model of car ownership and private car use. *European Economic Review* **34**:971-985

Knight F H (1924): Some fallacies in the interpretation of social cost. *Quarterly Journal of Economics* **38**:582-606.

Kulmala R (1995): Safety at rural three- and four-arm junctions. Development and application of accident prediction models. VTT Publications 233, Technical Research Centre of Finland, Espoo

Liem T, Dagenais M & Gaudry M (1993): LEVEL: the L-1.4 program for BC-GAUHESEQ regression - Box-Cox Generalized AUtoregressive HEteroskedastic Single EQuation models. Publication 510, Centre de recherche sur les transports, Université de Montréal.

Maddison D, Pearce D, Johansson O, Calthrop E, Litman T & Verhoef E (1996): *The true costs of transport*. Earthscan Publications Ltd, London.

McCarthy P (1999): TRAVAL-1: A Model for California. In: Gaudry & Lassarre (1999): *Structural Road Accident Models: The International DRAG Family*. Elsevier (forthcoming)

McCullagh P & Nelder J (1983): Generalized linear models. Chapman and Hall, New York

McFadden D (1974): Conditional logit analysis of qualitative choice behavior. Pp 105-142 in Zarembka P (ed): *Frontiers in econometrics*. Academic Press, New York.

McFadden D (1975): The revealed preferences of a government bureaucracy: theory. *Bell Journal of Economics* **6**(2):401-416.

McFadden D (1976): The revealed preferences of a government bureaucracy: empirical evidence. *Bell Journal of Economics* **7**(1):55-72.

McFadden D (1978): Modelling the choice of residential location. Pp 75-96 in Karlquist A et al (eds): *Spatial interaction theory and residential location*. North-Holland, Amsterdam.

McFadden D (1981): Econometric models of probabilistic choice. Pp 198-272 in Manski C F & McFadden D (eds): *Structural analysis of discrete data with econometric applications*. MIT Press, Cambridge, Mass.

Nash C (1997): Transport externalities: does monetary valuation make sense? Pp 232-254 in: Rus G de & Nash C (eds): *Recent developments in transport economics*. Ashgate Publishing Ltd, Aldershot.

Nelder J A & Wedderburn R W M (1972): Generalized linear models. *Journal of the Statistical Society A* **135**:370-384.

Newbery D (1988): Road user charges in Britain. *Economic Journal* **98** (Conference 1988):161-176.

Nilsson J-E (1991): Investment decisions in a public bureaucracy: A case study of Swedish road planning practices. *Journal of Transport Economics and Policy* **25**:163-175

Nyborg K & Spangen I (1996): Politiske beslutninger om investeringer i veger. Working Report 1026, Institute of Transport Economics, Oslo

Odeck J (1991): Om nytte-kostnadsanalysenes plass i beslutningsprosessen i vegsektoren. *Sosialøkonomen* no 3:10-15

Odeck J (1996): Ranking of regional road investment in Norway: Does socioeconomic analysis matter? *Transportation* **23**:123-140

Oppe S (1989): Macroscopic models for traffic and traffic safety. *Accident Analysis & Prevention* **21**:225-232.

Oppe S (1991a): The development of traffic and traffic safety in six developed countries. *Accident Analysis & Prevention* **23**:401-412.

Oppe S (1991b): Development of traffic and traffic safety: Global trends and incidental fluctuations. *Accident Analysis & Prevention* **23**:413-422.

Oum T H, Waters W G & Yong J-S (1992): Concepts of price elasticities of transport demand and recent empirical estimates. *Journal of Transport Economics and Policy* **26**:139-154.

Partyka S C (1984): Simple models of fatality trends using employment and population data. *Accident Analysis & Prevention* **16**:211-222.

Partyka S C (1991): Simple models of fatality trends revisited seven years later. *Accident Analysis & Prevention* **23**:423-430.

Peltzman S C (1975): The effects of automobile safety regulation. *Journal of Political Economy* **83**:677-725.

Pigou A C (1920): *The economics of welfare*. Macmillan, London.

Poisson S D (1837): *Recherches sur la probabilité des jugements en matière criminelle et en matière civile, précédées des règles générales du calcul des probabilités.* Bachelier, Paris.

Poisson S D (1841): *Lehrbuch der Wahrscheinlichkeitsrechnung und deren wichtigsten Anwendungen.* Meyer, Braunschweig.

Ramjerdi F & Rand L (1992): *The national model system for private travel.* Report 150, Institute of Transport Economics, Oslo.

Recht J L (1965): Multiple regression study of the effects of safety activities on the traffic accident problem. National Safety Council, Chicago.

Robertson L S (1981): Automobile safety regulation and death reductions in the United States. *American Journal of Public Health* **71**:818-822.

Salmon W C (1984): *Scientific explanation and the causal structure of the world.* Princeton University Press, Princeton.

Satterthwaite S P 81981): A survey of research into relationships between traffic accidents and traffic volumes. Supplementary Report 692, Transport and Road Research Laboratory, Crowthorne.

Smeed R J (1949): Some statistical aspects of road safety research. *Journal of the Royal Statistical Society A* **112**:1-34.

Smeed R J (1955): Accident rates. *International Road Safety & Traffic Review*, **3**:30-40.

Smeed R J (1974): The frequency of road accidents. *Zeitschrift für Verkehrssicherheit*, **20**:95-108 and 151-159.

Tegnér G & Loncar-Lucassi V (1996): Tidsseriemodeller över trafik- och olycksutvecklingen. Transek AB, Stockholm.

Theil H (1969): A multinomial extension of the linear logit model. *International Economic Review* **10**:251-259.

Train K (1986): *Qualitative choice analysis: Theory, econometrics and an application to automobile demand.* MIT Press, Cambridge, Mass.

Verhoef E (1996): *The economics of regulating road transport.* Edward Elgar, Cheltenham.

Wilde G J S (1972): General survey of efficiency and effectiveness of road safety campaigns: achievements and challenges. *Proceedings of the International Congress on the Occasion of the 40th Anniversary of the Dutch Road Safety Association,* The Hague.

Wilde G J S (1975): Road user behaviour and traffic safety: toward a rational strategy of accident prevention. Paper presented at the Annual Convention of the Dutch Road Safety League, Amsterdam..

Wilde G J S (1982): The theory of risk homeostasis: Implications for safety and health. *Risk Analysis* **2**(4):209-225.

Williams H C W L (1977): On the formation of travel demand models and economic evaluation measures of user benefit. *Environment and Planning* **9**:285-344.

Winkelmann R & Zimmermann K F (1992): Recursive probability estimators for count data. Münchener Wirtschaftwissenschaftliche Beiträge 92-04, Volkswirtschaftliche Fakultät, Ludwig-Maximilians-Universität, München. Presented at the EC^2 conference «Econométrie des Modèles de Durée, de Comptage et de Transition», Paris, December 10-11, 1992.

Zlatoper T J (1984): Regression analysis of time series data on motor vehicle deaths in the United States. *Journal of Transport Economics and Policy* **18**:263-274.

Zlatoper T J (1987): Factors affecting motor vehicle deaths in the USA: some cross-sectional evidence. *Applied Economics* **19**:753-761.

Zlatoper T J (1989): Models explaining motor vehicle death rates in the United States. *Accident Analysis & Prevention* **21**:125-154.

Zlatoper T J (1991): Determinants of motor vehicle deaths in the United States: a cross-sectional analysis. *Accident Analysis & Prevention* **23**:431-436.

Essay 1

The barely revealed preference behind road investment priorities

Public Choice **92:** 145–168, 1997.

The barely revealed preference behind road investment priorities*

LASSE FRIDSTRØM & RUNE ELVIK
Institute of Transport Economics, P.O. Box 6110 Etterstad, N-0602 Oslo, Norway

Accepted 15 July 1996

Abstract. An attempt is made to reveal the preference of decision makers within the regional Norwegian public roads administration. The order of priority assigned to the respective, competing public road investment opportunities within the various counties (provinces) is studied by means of a rank order multinomial logit model. Explanatory variables used include cost, benefit, and a variety of attributes characterizing the individual investment projects. Although statistically significant, cost and benefit appear to be of only marginal importance for the priorities set. More weight is attached to cost than to benefit. Smaller projects are preferred to larger, given the benefit-cost ratio. In general, the models estimated are able to explain only a relatively small share of the priority setting made.

Key words: public roads, investment, revealed preference, cost-benefit, logit model

1. Introduction

The decision process governing the allocation of resources between competing road investment projects has been the subject of several recent studies in Scandinavia (Elvik, 1993; Odeck, 1991 and 1996; Jansson and Nilsson, 1989; Nilsson, 1991; Nyborg and Spangen, 1996). In general, these studies have shown very weak – if any – association between the priority ranking assigned to a given road investment project and the project's calculated economic cost, benefit, or benefit/cost ratio. This may seem surprising in view of the fact that, if a maximum economic benefit is to be obtained within the constraint of a given investment budget, an optimal decision rule would be to rank the project according to a decreasing benefit/cost ratio, and then carry out the projects in that order, until the budget is depleted.

In this paper, we set out to reexamine this issue by means of potentially more powerful statistical methods than previously adopted. When previous studies have been able to detect virtually no association between benefit/cost

* Thanks are due to James Odeck and Toril Presttun for their advice and help in making the data set available, to Jan Erik Lindjord and Lasse Torgersen for their file management assistance, to Erik Biørn and Inger Spangen for their valuable comments, and to the PROSAF program of the Royal Norwegian Council for Scientific and Industrial Research for supporting the analysis financially.

ratio and priority ranking, could the reason be that these studies fail to take into account ("control for") certain fairly important constraints to which decision makers are subject? Will a different picture emerge through an appropriate, multivariate method of analysis, in which one estimates the *partial* effect of economic cost and benefit *conditional* on the relevant constraints? Are there a lot of qualitative factors at play, beyond those taken account of in the benefit/cost calculations, that influence decision making? If so, is it possible to identify these factors?

In short, we would like to be able to *reveal* the public decision makers' preference, in much the same way as in the pioneering papers by McFadden (1975, 1976). To this purpose, we have analyzed the data included in the Norwegian National Road Plan (NRP) for the period 1990–93, using, first, a rank order logit model for the detailed ranking within the respective counties, and, second, a traditional four-alternative logit model explaining crude priority assignments for all the projects considered. By contrasting these two lines of analysis we attempt to shed light on the possibility of a certain strategic (budget maximizing) behavior on the part of the regional road agencies.

2. Road investment planning in Norway

Public roads in Norway are divided into three administrative classes: national roads, county roads, and municipal roads. National roads are owned by the state. Investment planning for national roads is conducted by the Public Roads Administration, which consists of a central agency located in the capital and 19 regional offices, one in each county (province). The planning is done in the form of four-year investment plans – the NRPs.

The planning process starts two to three years before the final plan is to be presented. Thus, the National Road Plan for 1990–93 was drawn up during the years 1987–89. The central agency of the Public Roads Administration sets a timetable for planning and issues guidelines and instructions to its regional offices. Each regional office of the Public Roads Administration is required to conduct cost-benefit analyses of investment projects and to make a formal list of priorities among these projects. Plans are made by each of these offices and are henceforth sent to the central agency for review. They are then put together to form the National Road Plan (NRP), which is, in turn, submitted to the Parliament for approval.

3. A data set taken from the National Road Plan 1990–93

For the 1990–93 planning term, the central agency of the Public Roads Administration created an investment project data base containing all projects ranked and listed formally by the regional road agencies. More than 700 projects were included in this data base. Projects located in the county of Oslo (the capital) were, however, not included. In addition, the county of Aust-Agder had to be omitted for the purpose of our analyses, since the benefit-cost ratio had been defined in a way differing from that of other counties.

In addition to investment projects, the list of priorities made for each county contained a few "projects" that were not really investments, such as planning and transfer payments to municipalities. Projects of this kind were also omitted for the purpose of our analyses. We were thus left with a data base consisting of 686 investment opportunities located in 17 different counties.

An overview of the variables used in our study is given in Table 1. Most variables have been taken directly from the NRP data base, while a few have been constructed or recoded on the basis of combined information from the NRP and elsewhere. The MUNREP variable, for instance, has been constructed through manual inspection of the respective county ranking lists, along with maps facilitating the exact location of proposed road works.

It should be noted that we do not study the relative priorities between projects in different counties – only the internal ranking *within* each county is subject to analysis. The number of projects evaluated in the respective counties varies from 8 to 99.

4. A model for road investment choice behavior

Our objective is to understand and explain the assignment of priorities to the various road investment opportunities competing for funds within a given county of Norway. We shall do so by assuming that the regional decision-making unit can be pictured as behaving according to the logit model of discrete choice. A brief description of this modeling framework is therefore in order.

4.1. *The logit model of choice*

According to the standard, multinomial logit choice model, the probability that a decision maker n will choose alternative i from a choice set C_n can be written as follows:

$$p_n(i) = \frac{e^{V_{in}}}{\sum_{j \in C_n} e^{V_{jn}}} \tag{1}$$

where

Table 1. Independent variables

COST	Present value of investment costs (10^5 kr)
BENEFIT	Present value of benefits (10^5 kr)
BCRATIO	Benefit/cost ratio (10^{-2} – zero if missing)
BCRATIOZERO	1 if BCRATIO = 0, otherwise 0
USRBEN	Present value of user benefits (10^5 kr), consisting of (i) reduced vehicle operating costs, (ii) travel time savings, and (iii) benefit due to improved axle load tolerance
USBENZERO	1 if USRBEN = 0, otherwise 0
OWNBEN	Present value of road owner benefits (10^5 kr), consisting of (i) reduced road maintenance costs and (ii) reduced ferry costs
OWNBENZERO	1 of OWNBEN = 0, otherwise 0
ACCBEN	Present value of safety benefits (10^5 kr), i.e. reduced accident costs
ACCBENZERO	1 if ACCBEN = 0, otherwise 0
NOISBEN	Present value of noise abatement benefits
ACCIDENTS	Expected change in the number of accidents (10^{-1})
NOISEHOUSES	Number of dwellings affected by reduced noise
TIMEGAIN	Travel time gain per road user (minutes)
BLACKSPOTS	Accident blackspots treated, as referred to the 1983–86 record
AXLOADKMS	Number of road kilometers on which axle load tolerance is increased to 10 tons
AADT	Predicted annual average daily traffic as of the year 2000
MOTORWAY	1 if motorway, 0 otherwise
SCNDRYRD	1 if secondary road, 0 otherwise (i.e., if principal road)
MAINROAD	1 if main long distance itinerary road, 0 otherwise
CITY	1 if in major city area, 0 otherwise
NEWROAD	1 if new road, 0 otherwise
BIKEWALK	1 if walk or bicycle path, 0 otherwise
SAFETY	1 if special safety measure, 0 otherwise
PUBLIC	1 if measure to improve public transit, 0 otherwise
PLANREQ	1 if a local development plan is legally required, however not yet in existence, 0 otherwise
NOPLANREQ	1 if no development plan required, 0 otherwise
KMSFASTER	Number of road kilometers on which the speed can be increased from less than 80 kms/h to at least 80 kms/h
EARLY	1 if construction work is planned to start in 1989 or earlier, 0 otherwise
STARTED	1 if construction has already started and must be continued in order for previous investments to pay off, 0 otherwise
DECIDED	1 if project has been approved by Parliament, but not yet started, 0 otherwise
MUNREPZERO	1 if the host municipality in question is not already represented through a higher ranked project, 0 otherwise
MUNREP	Number of times the host municipality in question is already represented through higher ranked projects + MUNREPZERO
DPLAN	1 if a detailed area plan has been approved, 0 otherwise
LOCPLAN	1 if a local development plan has been approved, 0 otherwise
MASTERPLAN	1 if a road master plan has been approved, 0 otherwise
NOPLAN	1 if no development plan has been approved, 0 otherwise

$$U_{jn} = V_{jn} + w_{jn} \tag{2}$$

and

$$V_{jn} = \sum_{k=1}^{K} \eta_k x_{kn}(j). \tag{3}$$

Here, U_{jn} is the "indirect utility" of alternative j, consisting of a systematic part V_{jn} and a random part w_{jn}. The systematic part depends on a parameter vector $\eta = (\eta_1, ..., \eta_K)'$ and a vector of independent variables $\mathbf{x}_n(j) = (x_{1n}(j), ..., x_{Kn}(j))'$. It can be shown (see, e.g., Ben-Akiva and Lerman, 1985) that if the random terms w_{jn} are mutually independent and follow the Gumbel (double exponential) distribution, then a decision maker maximizing expected utility will choose alternative i with a probability given by equation (1).

The variables entering the logit indirect utility function may be of almost any kind, as long as the utility function is *linear in the parameters* $\eta_1,...,$ η_K and there is sufficient variability in the data set to make all parameters identifiable.[1]

4.2. *The rank order logit model*

In the Norwegian National Road Plan, for each county a unique order of priority is defined between the various projects, ranging from 1 to r(f) – the number of projects evaluated in county f.

To exploit the information contained in this ranking, we may infer – for each county f – a set of r(f)–1 implicit choices: project 1 is preferred to projects 2, 3,..., r(f), project 2 is preferred to projects 3, 4,..., r(f), and so on until the choice of project r(f)–1 versus r(f). In the first case, the choice set (C_1) consists of all r(f) projects evaluated in county f. In the second case, project 2 is considered chosen, while project 1 is considered unavailable, the choice set (C_2) consisting of the last r(f)–1 projects. In the last case, the choice set ($C_{r(f)-1}$) consists of the two projects ranked r(f)–1 and r(f).

In this way, the mutual ranking of r(f) projects if "exploded" into r(f)–1 implicit choice situations. It has been shown (Chapman and Staelin, 1982) that, under standard assumptions concerning the underlying utility functions (independent and identically distributed random utilities following the double exponential distribution), the r(f)–1 exploded choices can be treated as statistically independent outcomes of a conditional multinomial logit probability process. The factors governing these choices can thus be analyzed by means of standard logit modeling techniques, provided due account is taken of the gradually more restricted choice sets to be applied to the lower[2] ranked projects.

Given, however, that the maximum number of projects ranked in any one county is 99, yielding 98 implicit exploded choices among up to 99 alternatives, it has been necessary to reduce and simplify the data for the purpose of efficient estimation. We have done so by applying the method of *alternative sampling*. For each project (except the last one in each county), a simple random sample of (up to) ten lower ranked projects were sampled, forming a choice set of (up to) eleven alternatives (projects), of which the first one is always considered chosen. When there are less than ten lower ranked projects left on the county list, all of these are "sampled", forming a choice set consisting of two to ten projects. It has been shown (McFadden, 1978) that this procedure yields statistically consistent estimates of the logit model parameters. When, as in our case, the method of sampling alternatives is simple random sampling, no correction factor is necessary in the logit model in order to take account of the fact that the non-chosen alternatives are merely a sample from a larger available set (Ben-Akiva and Lerman, 1985).

In this way our data set is reduced to 669 observations on choices among (up to) eleven competing road investment opportunities. We analyze these data by means of the ALOGIT software package.

5. Empirical results

In Table 2, two different rank-order logit models are shown. The "ample model" accommodates a fairly large set of independent variables. In the "parsimonious model", only the significant coefficients from the ample model have – by and large – been retained.

5.1. *Impact of cost, benefit and project size*

One of the more obvious hypotheses to test is that the priority ranking assigned is determined in large part by the costs incurred and benefits derived from the project, or simply by their ratio.

There are, however, several conceivable ways to formulate the relationship between "utility" and cost/benefit, giving rise to varying economic interpretations. Having experimented with various forms, we base our analysis on the following, best-fit "logarithmic ratio model".

Let the vector η of equation (3) be partitioned as follows:

$$\eta' = (\kappa\, \beta'\, \gamma') \tag{4}$$

Table 2. Rank order logit models of within-county road investment priorities. Coefficient estimates, with standard errors in parentheses

Coefficient no.	Independent variable	Alt.[1]	Ample model	Parsimonious model
1	ln(COST)	GE	**–0.1659** (0.0432)	**–0.1592** (0.0366)
2	ln(BENEFIT/COST)	GE	0.08344 (0.0631)	**0.1375** (0.0423)
3	BCRATIOZERO	GE	–0.2455 (0.353)	
4	ln(USRBEN/BENEFIT)	GE	0.0940 (0.0574)	**0.1397** (0.0414)
5	USRBENZERO	GE	–0.3436 (0.243)	
6	ln(OWNBEN/BENEFIT)	GE	–0.0596 (0.0417)	*–0.0546* (0.0290)
7	OWNBENZERO	GE	–0.0996 (0.160)	
8	ln(ACCBEN/BENEFIT)	GE	*0.0866* (0.0442)	0.0367 (0.0299)
9	ACCBENZERO	GE	0.2304 (0.192)	
10	ln(MUNREP)	GE	**–0.4596** (0.0748)	**–0.4498** (0.0721)
11	MUNREPZERO	GE	0.0131 (0.0997)	0.0273 (0.0982)
12	EARLY	GE	**0.8753** (0.175)	**0.8710** (0.172)
13	STARTED	GE	**1.388** (0.194)	**1.367** (0.191)
14	DECIDED	GE	**1.371** (0.408)	**1.222** (0.397)
15	PLANREQ	GE	**–0.3246** (0.100)	**–0.3339** (0.0994)
16	MAINROAD	GE	**0.3698** (0.110)	**0.3486** (0.104)
17	CITY	GE	0.1960 (0.167)	
18	BIKEWALK	GE	**0.7126** (0.227)	**0.5609** (0.194)
19	SAFETY	GE	*0.5767* (0.254)	*0.4366* (0.235)
20	MOTORWAY	GE	–0.0667 (0.203)	
21	SCNDRYRD	GE	0.3975 (0.290)	
22	NEWROAD	GE	0.0380 (0.207)	
23	PUBLIC	GE	1.522 (1.29)	
	Log-likelihood		–1368.73	–1372.60
	ρ^2		0.0846	0.0820
	Units of observation		669	669
	Parameters		23	14

Italics: Significantly different from zero at the 10 per cent level by a two-sided test.
Bold face: Significantly different from zero at the 1 per cent level by a two-sided test.

Here, κ is thought of as the cost coefficient, β as a subvector of benefit coefficients, and γ as a coefficient subvector covering all other aspects taken into account in the decision process.

By the "logarithmic ratio model" we have in mind the following general formulation:

$$V_{kn} = \kappa \cdot \ln(\mathrm{COST}_n(k)) + \beta \cdot \ln\left(\frac{\mathrm{BENEFIT}_n(k)}{\mathrm{COST}_n(k)}\right) + \sum_j \gamma_j \cdot x_{jn}(k). \quad (5)$$

In simplified notation (suppressing subscripts and indices):

$$V = \kappa \cdot \ln(\mathrm{COST}) + \beta \cdot \ln(\mathrm{BENEFIT/COST}) + \sum_j \gamma_j \cdot x_j. \quad (6)$$

Does it matter to the decision maker whether the benefit is derived from, say, travel time gains, reduced road maintenance expenditure, a lower accident toll, or reduced noise? To address this question, we end up estimating a model in which the gross benefit is subdivided into user benefit, owner benefit, accident benefit and (the residual) noise-abatement benefit, as follows:

$$\begin{aligned} V = {} & \kappa \cdot \ln(\mathrm{COST}) + \beta_0 \cdot \ln(\mathrm{BENEFIT/COST}) \\ & + \beta_1 \cdot \ln(\mathrm{USRBEN/BENEFIT}) + \beta_2 \cdot \ln(\mathrm{OWNBEN/BENEFIT}) \\ & + \beta_3 \cdot \ln(\mathrm{ACCBEN/BENEFIT}) + \sum_j \gamma_j \cdot x_j \end{aligned} \quad (7)$$

This formulation gives rise to a number of useful interpretations. Since κ is the partial effect of cost, *given the benefit/cost ratio*, it can be interpreted as the effect of project *size*, i.e., the effect of *benefit and cost* increasing proportionately. The effect is estimated at $\hat{\kappa} = -0.1592$ (in the parsimonious model), meaning that the log-odds of preferring a given project (i, say) to some other project (j) is *reduced* by 0.1592 per cent for each per cent increase in the *size* of project i (the *odds elasticity*[3]). Other things being equal, small projects are preferred to large ones.

The effect of increased overall benefit, *given constant cost*, is estimated at $\hat{\beta}_0 = 0.1375$, with a 95 per cent approximate confidence interval ranging from 0.0546 to 0.2204.

As expected, increased payoff does tend to enhance a project's priority. But the effect is astonishingly small. To see this, consider two projects i and j that are initially identical, each having a 50 per cent probability of being preferred to the other. Now, suppose the benefit of project i is somehow doubled, while all the other characteristics (including costs) of projects i and j remain constant. This would cause the probability of preferring project i to j to increase by a mere 2.4 per cent[4] (confidence interval from 0.9 to 3.8 per cent). In other words, there would still be a 47.6 per cent probability for the decision maker to prefer a project providing only half as much benefit as the other, otherwise identical investment opportunity. Even a tenfold increase in benefit would, similarly, increase the choice probability by no more than 7.8 percentage points.

Now, imagine a project whose investment *cost* increases, while all other attributes (including benefit) are kept constant. The relevant odds elasticity is calculable as $\hat{\kappa}-\hat{\beta}_0 = -0.2967$ (confidence interval from -0.424 to -0.170). This amounts to saying that, if the cost is doubled, while the benefit is kept constant, the probability of preferring this project to another, equally attractive one is reduced by 5.1 percentage points (confidence limits -7.3 to -2.9 percentage points).

The fraction $(\hat{\beta}_0/(\hat{\kappa}-\hat{\beta}_0) = -0.463$ (confidence interval[5] from -0.617 to -0.310) has an interpretation as the marginal rate of substitution between a (say) one per cent increase in benefit and a one per cent increase in cost.[6] The analysis reveals, in other words, that in the mind of the decision maker, a – say – ten per cent increase in benefit is "worth" no more than a 4.6 per cent decrease in cost. Cost is, in a sense, about twice as important to the decision maker as benefit.[7]

It is, however, fair to say that neither cost nor benefit seems to have any major influence on decision making. It is striking to note that a (say) 10 per cent increase in *benefit* (given cost) has no bigger effect on priorities than a similar decrease in project *size*, i.e., if *both* cost and benefit are scaled down by 10 per cent.

The above estimates of the impact of project benefit on the priority ranking apply in the case where all benefit components are assumed to change proportionately. However, it does seem to matter to the decision maker whether benefit increases as a result of higher road user surplus, growing road owner benefit, or a reduction in accident cost. In general, road user benefit is valued at a higher rate than safety or noise reduction benefits, which in turn are given priority over road owner benefits. These differences are statistically significant, as witnessed by the estimates of coefficients β_1, β_2, and β_3 (referring to equation 7).

The corresponding odds elasticities,[8] expressing the effects of changes in USRBEN, OWNBEN, ACCBEN, and NOISBEN, are shown in Table 3. For each per cent increase in user benefit, the odds of preferring that particular project increases by an estimated 0.1506 per cent, given that cost and *other* benefit components are unchanged.[9] The estimated marginal rate of substitution with respect to a one per cent increase in cost is -0.508.

More surprisingly, the effect of road owner benefit on ranking has the counterintuitive sign. A reduction in road maintenance and ferry costs appears to *negatively* affect priority ranking, other things being equal. This effect is statistically significant at the 5 per cent level.

As for safety and noise reduction benefits, the effects are not significant.

Table 3. Odds elasticities with respect to various benefit components, as evaluated at overall benefit component shares

Benefit component	Point estimate	95 per cent confidence interval	
		Lower limit	Upper limit
User benefit (USRBEN)	**0.1506**	0.0687	0.2326
Road owner benefit (OWNBEN)	*−0.0526*	−0.1047	−0.0005
Safety benefit (ACCBEN)	0.0392	−0.0128	0.0912
Noise abatement benefit (NOISBEN)	0.0003	−0.0019	0.0025

Italics: Significantly different from zero at the 10 per cent level by a two-sided test.
Bold face: Significantly different from zero at the 1 per cent level by a two-sided test.

5.2. *Impact of non-economic factors*

In addition to cost and benefit, we have included in the model a successively larger set of "auxiliary variables" (i.e., the x_js), to account for various, potentially important considerations in the decision making process.

The simplest models, in which only cost, overall benefit, and a dummy for zero benefit-cost ratio are included, explain a very small fraction of the total variation (less than 0.6 per cent, as measured by the ρ^2 measure[10]). Only the cost term is statistically different from zero. When the benefit is decomposed, only insignificant improvements in fit are seen.

Interestingly, however, as we add more "auxiliary" variables to the model, the (absolute) estimated values and the statistical significance of the cost and benefit coefficients tend to increase. Thus, far from removing explanatory power from the cost and benefit factors, the addition of auxiliary variables appears necessary in order for the significance of cost and benefit to be revealed. When the auxiliary factors are not controlled for, however, the role of cost and benefit is much harder to spot.

5.2.1. *Geographic distribution*

The introduction of a dummy variable (MUNREPZERO) capturing whether the municipality concerned by a road project has already been represented ("favored") by a higher ranked project, improves the fit considerably. In such cases, the chances are reduced for a second (or third, fourth,...) project located in the same muncipality to receive a high priority, other things being equal.

An even stronger improvement in fit occurs when we enter into the model the (log of the) number of times the same municipality is already represented on the list of priorities (MUNREP). In these models, the MUNREPZERO term becomes insignificant, meaning that there is no important difference between zero and one previous representation. However, as the same municipality is represented more than twice, the decision maker appears increasingly concerned not to allocate an excessive amount of investment funds to one and the same geographical area.[11]

The coefficient of ln(MUNREP) indicates that, as one doubles the number of times a given municipality has so far been "favored" by a road project (say, from 1 to 2, or from 2 to 4), the odds of allocating still another project to that municipality are lowered by some 27 per cent (from, say, 1 to $2^{-0.4498} = 0.732$). This corresponds to an almost 8 percentage point decrease in probability, starting from an initial 50 per cent.

To produce an equivalent deterioration in odds through a cost alteration, the cost would have to increase almost three-fold. Alternatively, the total benefit of the project would have to be cut by 90 per cent.

5.2.2. *Project type*

Projects considered part of a main long distance itinerary road are given enhanced priority, equivalent to an estimated 69 per cent cost reduction or a more than 12-fold increase in benefit. Bicycle and walking paths are even more popular with the decision makers, receiving a strengthened priority equivalent to an 85 per cent cost cut or an almost 60-fold increase in benefit. Projects classified as traffic safety measures also appear to receive strongly enhanced priority, although this effect is just barely significant at the 5 per cent level.

The following factors were found *not* to be significant (confer Table 1): CITY, MOTORWAY, SCNDRYRD, NEWROAD, PUBLIC.[12]

5.2.3. *Planning requirements*

Under Norwegian law, all land development and building activities are subject to approval through detailed, local development plans or more general master plans, as set out in the Planning and Building Act of 1985. For instance, in order for a public authority to expropriate or acquire right of way across private property, a local development plan approved by the municipality involved is mandatory.

The NRP contains information as to the planning provisions required and the planning documents available for each individual project. It turns out that the existence of local development plans or road master plans *per se* does not have a significant impact on ranking. However, if a local development plan is indeed required, but lacking (PLANREQ = 1), then the odds of that particular project are significantly reduced, by a rate corresponding to a more than three-fold increase in cost, or a 91 per cent reduction in benefit. In essence, when the required legal basis for a road investment is lacking, the project will – not unexpectedly – in all likelihood be assigned a lower priority.

5.2.4. *Decision-making constraints*

In some cases, the priority ranking between competing road investment projects has to take account of certain constraints imposed, such as whether or not the project has previously been approved by Parliament (DECIDED), if it is scheduled to start already during (one of the) previous planning period(s) (EARLY), or if construction work has in fact already begun (STARTED).

The estimated "value" of these constraints is such as to overrule almost any difference in economic cost or benefit that might exist among the candidate projects. When, for instance, a project has already been approved by Parliament, its odds are improved by an estimated factor equivalent to 98 per cent cost cut or a 7 200-fold increase in benefit. Essentially, this means that a project not yet approved by Parliament will hardly ever stand a better chance than one which does enjoy previous approval, no matter how favorable the cost-benefit assessment comes out for the former.

In some cases, Parliamentary decisions concerning a road investment project have already been made, or construction may in fact have begun, in which case the project must be completed in order for previous investments to pay off. As expected, such instances tend to receive a relatively high priority. When the three variables EARLY, STARTED, and DECIDED are added to the model, the fit improves by more than 4 percentage points (in terms of ρ^2).

5.3. *Are cost-benefit analyses distrusted?*

Although the benefit and, especially, the cost terms do come out as at least marginally significant in models taking account of the more important constraints facing the decision maker, it is fair to say that the overall explanatory power of the models if relatively low. The confidence intervals around the benefit coefficient estimates are rather wide. The cost-benefit evaluation performed does not seem to have any large bearing on the priority ranking done. One reason for this might be that the decision maker, being aware of the errors, inaccuracies and weaknesses inherent in the cost-benefit analysis, deliberately chooses not to put too much weight on its outcome.

If it is true that the cost-benefit analysis lacks credibility with the decision maker, one might suspect that the objective characteristics of the investment project, as expressed *not* in monetary terms, but in their "natural" units of measurement, would perhaps be able to explain a larger share of decision-making behavior. In other words: do the crude input factors into the benefit evaluation carry more weight with the decision maker than the monetary benefit derived from them? To study this possibility, we estimate a set of "heuristic" ("non-economic") rank order logit models, in which the economic benefit terms have been replaced by their physical (engineering) counterparts,

Table 4. Heuristic, rank order logit models of within-county road investment priorities. Coefficient estimates, with standard errors in parentheses

Coefficient no.	Independent variable	Alt.[1]	Ample model	Parsimonious model
1	COST (100 mNOK)	GE	**–0.1803** (0.0671)	**–0.1846** (0.0619)
2	ACCIDENTS (1000)	GE	0.5666 (1.27)	0.9562 (1.14)
3	NOISEHOUSES (1000)	GE	–0.0553 (0.148)	–0.0343 (0.151)
4	AADT (million)	GE	3.1490 (8.16)	
5	TIMEGAIN (1000 min)	GE	-0.0552 (0.664)	
6	TIMEGAIN × AADT**	GE	*–0.4956* (0.288)	*–0.4663* (0.273)
7	AXLOADKMS (1000)	GE	0.2471 (0.882)	
8	AXLOADKMS × AADT*	GE	0.0147 (0.349)	0.0602 (0.288)
9	KMSFASTER (1000)	GE	–0.8967 (2.69)	
10	KMSFASTER × AADT*	GE	0.1589 (0.236)	
11	MUNREPZERO	GE	**0.2777** (0.0913)	**0.2773** (0.0911)
12	EARLY	GE	**0.9280** (0.177)	**0.9227** (0.175)
13	STARTED	GE	**1.203** (0.196)	**1.216** (0.192)
14	DECIDED	GE	**1.461** (0.447)	**1.485** (0.444)
15	PLANREQ	GE	**–0.3263** (0.0980)	**–0.3283** (0.0979)
16	MAINROAD	GE	0.1666 (0.103)	0.1799 (0.0993)
17	CITY	GE	–0.0246 (0.175)	–0.0079 (0.159)
18	BIKEWALK	GE	**0.5609** (0.183)	**0.5638** (0.182)
19	SAFETY	GE	*0.4762* (0.205)	*0.4835* (0.206)
20	BLACKSPOTS	GE	*0.1630* (0.080)	*0.1648* (0.079)
	Log-likelihood		–1401.44	–1372.60
	ρ^2		0.0627	0.0820
	Units of observation		669	669
	Parameters		20	15

Italics: Significantly different from zero at the 10 per cent level by a two-sided test.
Bold face: Significantly different from zero at the 1 per cent level by a two-sided test.
* Million vehicle kilometers.
** Million vehicle minutes.

such as the travel time gain per road user (minutes), the expected traffic volume (vehicles per day), the expected number of accidents avoided, the number of houses less affected by noise, the number of road kilometers with increased axle load tolerance, the number of road kilometers allowing for higher speed, the number of accident blackspots treated[13] etc. The results of this exercise are shown in Table 4.

A linear[14] cost term has been retained even in the "non-economic" models, together with the significant variables from the economic models. The substitution of crude, project attribute input factors for the calculated cost-benefit output does not, however, improve the model's explanatory power. With one exception, all the project attributes considered are statistically insignificant, and the one significant coefficient has the "wrong" sign. Our heuristic models are able to explain no more than 6 to 7 per cent of the variation in priority ranking.

6. Testing for strategic priority assignment

Prior to drawing up their list of priorities, each regional public roads agency receives a preliminary investment budget from the central agency. For the 1990–93 planning term, the preliminary budget of each regional office contained three alternative spending limits. One of these was designated as the main alternative, or the "100 per cent" spending limit. The other two were 20 per cent below and 20 per cent above the 100 per cent limit, respectively. These three alternative spending limits divide the investment projects into four classes: (1) "safe projects", i.e., projects that were assigned priority within 80 per cent of the main spending limit; (2) "marginal projects if spending is cut", i.e., projects ranked within the 80 to 100 per cent range of the main spending limit; (3) "marginal projects if spending is increased", i.e., projects ranked within the 100 to 120 per cent range, and (4) "less eligible projects", i.e., projects with a priority too weak to be comprised by 120 per cent spending limit.

Funding for marginal projects is, ideally, to be based strictly on their benefit-cost ratios. The central agency of the Public Roads Administration is expected to compare marginal projects from all counties and provide funding for those which pass the cost-benefit test. There is thus an incentive for each county to place some of its better projects in the marginal categories, in order to maximize the chances of obtaining funds in excess of the 80 per cent limit, during the final round of appropriations undertaken by the national authorities (in principle the Parliament). Elvik (1995) has demonstrated how the distribution of state funds for national road investments between Norwegian counties can be understood in terms of a vote trading ("log rolling") process.

We would like to test the tenability of such a hypothesis concerning regional decision-making behavior. In the following, we therefore set out to explain how projects are assigned to one out of the four spending limit classes. To this purpose, we estimate a traditional, four-alternative multinomial logit model, using each project contained in the NRP file as one observation.

As for the rank order logit models, we start out by estimating a fairly comprehensive, exploratory ("ample") model. This model contains 43 free parameters, all of which are alternative-specific[1], although in some cases the value is constrained to be equal between two or three alternatives.

Many of these parameters are, however, not significantly different from zero or from each other. By applying a suitable set of additional parameter constraints we are able to reduce the number of parameters to 16, without a statistically significant drop in explanatory power (p-value of 0.51 by the log-likelihood chi-square test). Hence we arrive at the "parsimonious" model. Both models are exhibited in Table 5.

For all coefficients, we define class 2 as our reference class, meaning that all variables with non-zero coefficients enter the "utility" functions of classes 1, 3 or 4, *expressing partial effects on the log-odds with respect to class 2*.

The size of the project (as measured by its COST) has a negative impact on the probability of being assigned to the top priority class (1), but a positive impact on the probability of classes 3 or 4, as compared to class 2. In the parsimonious model we constrain the coefficients assigned to classes 3 and 4 to be equal. The coefficient sign pattern is consistent, in that the larger the project is, the weaker priority it tends to receive. This result is well in line with the rank order model estimates.

With respect to the benefit-cost ratio, however, certain contradictions appear. There is no significant tendency to assign "good" projects (i.e., projects with a high benefit-cost ratio) to the top priority ("safe") class rather than to the second priority ("marginal") class, as one would expect from a decision maker aiming to maximize expected benefit within a given investment budget, and as one would also expect in view of the rank order model estimates. On the average, the projects of class 2 are neither more nor less profitable than those of class 1.

While, in the mutual ranking of projects, decision makers tend to assign somewhat enhanced priorities to the more profitable ones, it may appear as if this ordering takes place only *within* each priority class. A certain number of fairly good projects may seem to be kept in reserve for class 2. There is, however, a certain (significant) tendency for less socially profitable projects to end up in the weak priority classes (3 or 4).

The ample model reveals no statistically significant correspondence between type of benefit accruing and the priority class assignment. However, the sign pattern emerging is generally consistent with the findings of the rank order logit model, in that road user and safety benefits tend to improve the priority, while the opposite is true of road owner benefits. In the parsimonious model, all these coefficients are constrained to zero.

As in the rank order model, decision making constraints such as previous Parliamentary approval or onset of construction is seen to have a marked impact on priority class assignment, almost all projects subject to this kind of constraints being assigned to the "safe" priority class, and very few (if any) to classes 3 or 4.

Table 5. Logit models of priority class assignment. Coefficient estimates, with standard errors in parentheses

Var no.	Independent variable	Alt.[1]	Ample model	Parsimonious model
1	ln(COST)	1	**–0.2383** (0.0822)	**–0.1809** (0.0700)
2		3	*0.1665* (0.0983)	*0.1426* (0.0723)
3		4	0.0557 (0.113)	
4	ln(BENEFIT/COST)	1	–0.0686 (0.0892)	–0.0064 (0.0707)
5		3	*–0.2279* (0.109)	*–0.1796* (0.0750)
6		4	–0.2163 (0.134)	
7	ln(USRBEN/BENEFIT)	1	0.0285 (0.106)	
8		3	0.0377 (0.135)	
9		4	–0.1306 (0.150)	
10	ln(OWNBEN/BENEFIT)	1	–0.0821 (0.0680)	
11		3	–0.0885 (0.0791)	
12		4	0.0767 (0.114)	
13	ln(ACCBEN/BENEFIT)	1	–0.1017 (0.0775)	
14		3	–0.1000 (0.0925)	
15		4	–0.0956 (0.116)	
16	EARLY	1	**1.165** (0.421)	**1.172** (0.416)
17		3,4	*–1.678* (0.718)	*–1.586* (0.715)
18	STARTED	1	**1.612** (0.477)	**1.592** (0.474)
19		3,4	*–2.719* (1.13)	**–2.887** (1.12)
20	DECIDED	1	*1.436* (0.852)	*1.441* (0.840)
21	PLANREQ	1	*–0.4346* (0.243)	
22		3	*–0.6275* (0.313)	*– 0.4224* (0.229)
23		4	0.0521 (0.357)	
24	MAINROAD	1	*0.5427* (0.238)	
25		3	0.2880 (0.282)	0.3176 (0.207)
26		4	0.1255 (0.333)	
27	CITY	1	–0.2243 (0.341)	
28		3	*–0.7425* (0.434)	–0.2780 (0.285)
29		4	–0.3332 (0.607)	
30	BIKEWALK	1	*0.7305* (0.405)	0.5157 (0.371)
31		3	–1.003 (0.813)	*–1.221* (0.668)
32		4	–1.310 (1.10)	
33	SAFETY	1	*0.9770* (0.490)	
34		3	0.7829 (0.644)	*0.7358* (0.427)
35		4	0.9955 (0.627)	
36	MOTORWAY	1	0.4576 (0.506)	
37		3	0.8204 (0.544)	*0.8135* (0.455)
38		4	*1.420* (0.827)	
39	SCNDRYRD	1	0.4929 (0.585)	
40		3,4	–0.8338 (0.880)	

Table 5. Continued

Var no.	Independent variable	Alt.[1]	Ample model	Parsimonious model
41		1	−0.2569 (0.453)	
42	NEWROAD	3	−0.5555 (0.561)	
43		4	−0.7707 (0.647)	
44	ln(SIZEi)[a] (i=1,2,3,4)	GE	1 (by constraint)	1 (by constraint)
	Log-likelihood		−701.49	−714.55
	ρ^2		0.206	0.191
	Units of observation		686	686
	Parameters		43	16

Italics: Significantly different from zero at the 10 per cent level by a two-sided test.
Bold face: Significantly different from zero at the 1 per cent level by a two-sided test.
[a]The size of the indicated budget differs greatly between counties, and so does the monetary value of the projects proposed. Some counties propose no more projects that can be accommodated within the suggested budget – others list almost twice as many. To take account of the fact that the priority classes differ in size – for one and the same county as well as between counties – we include into the logit model a set of *size* variables, one for each "utility" function (not to be confounded with *project* size, as characterized by its cost and benefit). For instance, for a county having proposed projects corresponding to 112 per cent of the indicated budget, hence depleting class 1 and 2, but not class 3, the SIZE1 variable (entering the "utility" function of priority class 1) would equal 0.8, while SIZE2 = 0.2, and SIZE3 = 0.12. Alternative 4 has zero size and is defined as unavailable, having a zero probability of being "chosen". In the above example, therefore, if no other independent variables had explanatory power, or were included in the model, the probability of falling in class 1 would be 0.8/(0.8 + 0.2 + 0.12) = 0.714. This appears reasonable since, in this case, class 1 is large enough to accommodate exactly 71.4 per cent of the investments considered. If, in other words, projects were ranked through an entirely random procedure, the probability for a given project to end up in class 1 would be 71.4 per cent.

Interestingly, a number of variables appear to affect the assignment to classes 1, 3 and 4 in about the same way, meaning – in essence – that class 2 ("marginal projects if spending is cut") is the "odd man out" compared to the other classes. For instance, if a project is classified as a "traffic safety measure" (i.e., if the SAFETY dummy is one), the log-odds of being assigned to class 1 versus class 2 is improved (by 0.997 according the ample model). But so are the log-odds of classes 3 and 4 (with respect to class 2)! Indeed, the coefficient pertaining to classes 1, 3 and 4 are barely different, meaning that they differ from class 2 in about the same way. (Hence, they are constrained to be equal in the parsimonious model).

A similar pattern of effects is seen for other variables as well. The MAIN-ROAD, CITY, MOTORWAY, and PLANREQ criteria all exhibit effects that are not statistically different between alternatives 1, 3 and 4. They are, however, in most cases statistically different from zero, meaning that the probability of being assigned to class 2 is, indeed, affected.

In summary, there appears to be a tendency, among regional road authorities, to avoid the assignment of motorway, road safety, or main long distance itinerary road projects to priority class 2 ("marginal projects if spending is cut"). Projects for which the legally required area development plans are currently lacking, tend, however, to end up in this category. The same may be true of projects located to major urban areas, although this tendency is not statistically significant.

For projects characterized as bicycle and walking paths, the pattern of effects is distinctly different. These tend to be assigned to the "safe" priority class. For given size and benefit-cost ratio, bicycle and walking paths also have a lower than average risk of being placed in the weak priority classes (3 or 4). In other words, bicycle and walking paths are consistently given enhanced priority compared to other types of construction work.

7. Discussion

Why is it that our models are unable to explain a larger part of decision-making behavior? If the variables included in our analyses fail to explain the priorities, what does? Is the process of assigning county level priorities to competing road investment projects essentially a random one, with only a small systematic component?

In a recent interview survey among members of the Norwegian Parliament (MPs), the issues of road investment decisions and the use of cost-benefit analyses were raised (Nyborg and Spangen, 1996). Although their study focuses on the final, national Parliamentary decision making rather than on the foregoing regional planning procedure, to which our data refer, we believe their findings to have substantial relevance as a source of understanding even the local decision-making process. While the National Road Plan is being drawn up and prepared by the Public Roads Administration, the members of the Parliamentary Transport Committee visit every county, discuss with local politicians and administrators, and so form their own opinion on the projects being proposed. Also, the priority setting within each county is not an entirely bureaucratic process. The proposals made by the regional road agencies are submitted to and discussed by the County (provincial) Assembly, before being amended and passed on to the central authorities. Thus, the national political and local bureaucratic levels of planning are by no means disconnected. Local planners, being well acquainted with the political interests and pressures being exercised in one or the other direction, and having learned from experience what kinds of projects are likely to obtain political assent, adapt their list of priorities to the assumed politically acceptable or feasible.

Most members of Parliament accept the benefit-cost ratio as an important aspect of a road investment project, although none of them would think of it as the only relevant criterion. As several MPs would put it: If decisions were to be made based on the benefit-cost ratio alone, politicians would be redundant. Some MPs were outright negative to the use of cost-benefit analysis, seeing it as an attempt to quantify the unmeasurable.

Several MPs expressed concern over the fact that the user benefit was not decomposable between corporate and private road users. While many politicians would put considerable emphasis on time gains benefiting the local business community, they felt that leasure time gains among private individuals were a lot less tangible, and perhaps should not be valued (above zero) at all.[15]

Income distribution is an important issue among politicians. In Norway, this is particularly so when distribution is linked to the geographic dimension, as in the case of competing road investments. The MUNREP effect estimated by us is but one indication of this. More importantly, perhaps, many MPs stated that they would give special precedence to projects benefiting economically depressed areas.

Now, economic depression is likely to be *inversely* related to the user benefit derived from a local road investment, since the industry and population base, and hence the traffic volume to which time gains apply, would tend to be smaller in such areas than in the more prosperous ones. To the extent that distributional considerations are playing an important role even among local planners, it seems quite probable that this could explain the apparently low emphasis put on benefit during the ranking procedure. We do not, unfortunately, possess the information necessary to test such an hypothesis.

Several MPs characterize the road investment budget as "comfortable", a rather rare occurrence within the public sector. Again, if this perception is reflected even among local planners, it may help explain why cost does not come out as a more decisive factor. Since almost all the road investment projects evaluated by the public roads administration are sooner or later carried out (i.e., in the course of two or three consecutive planning periods), the incentive to screen projects in a highly conscientious and meticulous manner may appear to be relatively weak.

Other issues considered important by the MPs interviewed were: environmental effects (noise, scenery, water resources, recreation areas, nature conservation), public transportation level-of-service, school-children's safety, bicycle and pedestrian paths, main long distance itineraries, and rockslide prevention, to mention some of the most important issues. Some of these have been included in our model, others are rather judgemental or hard to formalize and hence have not been quantified in the National Road Plan.

Odeck (1996) administered a questionnaire to the regional road agencies, in an attempt to establish their ranking criteria by direct questioning. In descending order, the following criteria were found: (i) road safety, (ii) noise, (iii) user benefit, (iv) air pollution, (v) system continuity, (vi) regional economic development, (vii) local or national political wishes, (viii) benefit-cost ratio, (ix) seniority of proposal (from previous planning periods), and (x) relation to other modes of transport. This "stated preference" is in rather strong contradiction with our "revealed preference" estimates, which indicate a quite weak dependence on the three topmost criteria on the list.[16] Odeck (1996) remarks that criteria (i), (ii), (iii), (vi) and (viii) are all embedded in the benefit-cost ratio and that considering these in addition to the economic evaluation would imply double counting.

Odeck also asked the regional road agencies why they did not use the benefit-cost ratio in their ranking. He cites the following two most frequent answers: (a) that road investments should be used to foster economic development in depressed areas, and (b) that certain quite important (environmental) elements are not included in the cost-benefit evaluation.

One notes that these arguments are well in accordance with the views expressed by certain members of Parliament.[17]

8. Conclusions

Our study is concerned with the priority ranking between candidate, regional road investment opportunities in Norway. Using a (near) maximum of available information from the National Road Plan, we find, unlike previous studies, that economic cost and benefit do have a statistically significant impact on the ranking. In absolute terms, however, their impact is rather small, and in no way decisive. Cost is revealed to be about twice as important as benefit. Among the benefit components, road user benefits (reduced vehicle operating costs, travel time savings, and improved axle load tolerance) are the most decisive, while road owner savings in the form of reduced maintenance or ferry operating costs appear to have a surprising negative effect on ranking. Safety or noise abatement benefits have no significant effects.

There are indications that regional decision makers attempt to spread the opportunities fairly evenly among the municipalities belonging to the county. Chances are that no one municipality will be "favored" by more than a few projects, no matter how socially profitable its candidate projects may be. Also, there are strong indications, based on surveys within the bureaucratic as well as the political system, that income distribution concerns lead planners to favor projects benefiting economically depressed areas. This would have the likely (partial) effect of systematically promoting projects with weak profitability.

Confronting two different lines of analyzing the road investment priority settings, we find certain (weak) indications that regional road authorities do apply a kind of strategic priority assignment, allocating certain types of projects to the competitive ("marginal") priority class. There is no sign that economically stronger projects (i.e., projects with a favorable benefit-cost ratio) have a higher probability of being assigned to the marginal class. On the other hand – they do not have a lower probability of ending up in this category, either, as one would expect under (non-strategic) economic welfare maximization.

Notes

1. The variables of a multinomial logit model may be *generic* (GE) or *alternative specific.* A *generic* independent variable is a variable whose coefficient is constrained to be equal across all alternatives. In order for a generic coefficient (η_k, say) to be identifiable, the variable (x_{kn}) must exhibit some degree of variation *across alternatives*, i.e., we must have $x_{kn}(i) \neq x_{kn}(j)$ for some triplet (i,j,n). An *alternative-specific* variable enters the respective utility functions (alternatives 1,2,3,4,..., say) with unequal coefficients. Here, identifiability generally requires variation, not across alternatives, but *across the units of observation*, i.e., we must have $x_{km}(j) \neq x_{kn}(j)$ for some pair (m,n). (More precisely, for each alternative except one, it is possible to identify the coefficient of one and only one variable which does not vary across the sample – the alternative-specific *constant.*)
2. We refer to project j as "lower ranked" than project i if i < j, i or j being the priority assigned to the two projects. That is, we visualize a list in which priority 1 is on top, while the least attractive project (priority r(f)) is placed at the bottom.
3. In the logit choice model, the log-odds between any two alternatives is a linear function of the parameter vector η:

$$\ln[p_n(i)/p_n(j)] = V_{in} - V_{jn} = \sum_{k=1}^{K} \eta_k[x_{kn}(i) - x_{kn}(j)]. \tag{N1}$$

The (direct) *odds elasticity* with respect to a given attribute $x_{kn}(i)$ is therefore given by

$$\frac{\partial\left(\frac{p_n(i)}{p_n(j)}\right)}{\partial x_{kn}(i)} \Bigg/ \frac{\left(\frac{p_n(i)}{p_n(j)}\right)}{x_{kn}(i)} = \eta_k x_{kn}(i). \tag{N2}$$

If $x_{kn}(i)$ happens to be a logarithmic function ($x_{kn}(i) = \ln[y_n(i)]$, say), then we can write

$$\frac{\partial\left(\frac{p_n(i)}{p_n(j)}\right)}{\partial y_n(i)} \Bigg/ \frac{\left(\frac{p_n(i)}{p_n(j)}\right)}{y_n(i)} = \eta_k, \tag{N3}$$

i.e., the odds elasticity with respect to $y_n(i)$ is a constant and given by the corresponding logit coefficient.

4. If the two alternatives (i and j, say) are initially equally attractive, we have $V_{in} = V_{jn}$. Assume, in keeping with the notation of the previous note, that our variable $y_n(i) = e^{x_{kn}(i)}$ somehow increases by a factor c. The odds are then changed from 1:1 to c^{η_k}:1, the new conditional probability of choosing alternative i (given i *or* j) being calculable as

$$\frac{p_n(i)}{p_n(i)+p_n(j)} = \frac{c^{\eta_k}}{c^{\eta_k}+1}. \tag{N4}$$

The percentage point increase in the conditional probability, compared to the initial, 50 per cent probability, is then:

$$a(y, c:1) = 100 \cdot \left[\frac{c^{\eta_k}}{c^{\eta_k}+1} - \frac{1}{2}\right]. \tag{N5}$$

Putting c = 2 and $\eta_k = \hat{\beta}_0 = 0.1375$, we have a(y,2:1) = –2.4 in this case.

5. To calculate the variance of $\hat{\beta}_0/(\hat{\kappa}-\hat{\beta}_0)$ we use the Taylor approximation $\mathrm{var}(x/y) \approx \frac{1}{y^2}\left[\mathrm{var}(x) - 2\frac{x}{y}\mathrm{cov}(x,y) + \frac{x^2}{y^2}\mathrm{var}(y)\right]$, yielding, in the case $x = \hat{\beta}_0$ and $y = \hat{\kappa} - \hat{\beta}_0$,
$\mathrm{var}\left[\frac{\hat{\beta}_0}{\hat{\kappa}-\hat{\beta}_0}\right] \approx \frac{\mathrm{var}(\hat{\beta}_0)}{(\hat{\kappa}-\hat{\beta}_0)^2} - \frac{2\hat{\beta}_0\{\mathrm{cov}(\hat{\kappa},\hat{\beta}_0)-\mathrm{var}(\hat{\beta}_0)\}}{(\hat{\kappa}-\hat{\beta}_0)^3} + \frac{\hat{\beta}_0^2\{\mathrm{var}(\hat{\kappa})+\mathrm{var}(\hat{\beta}_0)-2\mathrm{cov}(\hat{\kappa},\hat{\beta}_0)\}}{(\hat{\kappa}-\hat{\beta}_0)^4}$

6. Under the assumption that decision makers are indeed maximizing some notion of utility, as formalized in the multinomial logic choice model (Section 4.1 above), one can derive marginal rates of substitution (MRS) between the various attributes entering the utility functions estimated:

$$\mathrm{MRS}[x_{hn}(j), x_{kn}(j)] \equiv \frac{\partial V_{jn}}{\partial x_{hn}(j)} \Big/ \frac{\partial V_{jn}}{\partial x_{kn}(j)} = \frac{\eta_h}{\eta_k} \equiv s_j(x_h, x_k)(\mathrm{say}). \tag{N6}$$

The decision maker trades a unit of attribute h against a unit of attribute k at a rate given by the ratio between their respective coefficients.

In the case of logarithmic variables, say $x_{hn}(i) = \ln[z_n(i)]$ and $x_{kn}(i) = \ln[y_n(i)]$, we obtain

$$s_j(z, y) = \frac{\eta_h}{\eta_k} \cdot \frac{y}{z}. \tag{N7}$$

Here, the MRS depends on the initial value taken on by either variable. It is, however, invariant as expressed in terms of relative (or percentage) changes. Suppose we compare a (say) c-fold change in z with a c-fold change in y (e.g., take c = 1.01 and compare one per cent changes). These two changes are traded against each other at a constant rate given by:

$$\frac{\eta_h \cdot [\ln(c \cdot z) - \ln(z)]}{\eta_k \cdot [\ln(c \cdot y) - \ln(y)]} = \frac{\eta_h \cdot \ln(c)}{\eta_k \cdot \ln(c)} = \frac{\eta_h}{\eta_k}. \tag{N8}$$

7. Since we are reasoning in terms of relative (percentage) changes, this rate of substitution does not necessarily mean that a dollar earned is worth less than half a dollar spent – it might simply mean that benefits are generally only half as large as cost. Such is, however, not the case. The sample mean benefit-cost ratio is about 1.4. In our alternative, "linear additive" formulation of our rank order logit model the dollar-to-dollar rate of substitution between benefit and cost comes out even smaller (in absolute value), viz. at –0.22.

8. These odds elasticities are given by the formula

$$\frac{\partial\left(\frac{p_n(i)}{p_n(j)}\right)}{\partial b_{ik}} \Big/ \frac{\left(\frac{p_n(i)}{p_n(j)}\right)}{b_{ik}} = \left[\beta_k + \frac{\left(\beta_0 - \sum_{i=1}^{3}\beta_i\right) b_{ik}}{b_i}\right] = \varepsilon_k(\mathrm{say}), (k = 1, 2, 3, 4) \tag{N9}$$

where b_{ik} is the k'th benefit component, b_i is the total benefit of project i, and β_k are the benefit coefficients as defined in the parsimonious rank order logit model (equation 7). By convention, $\beta_4 = 0$, the (residual) noise benefit coefficient. Note that the elasticity depends on the share b_{ik}/b_i of total benefit pertaining to the k'th benefit component. As a realistic numerical example, we use the overall sample totals $\sum_i b_{ik}/\sum_i b_i$ in our calculations. As summed through all the projects considered, user benefit represents 70 per cent, road owner benefit 13 per cent, accident reduction benefit 16 per cent and noise reduction benefit 2 per cent.

9. *Total* benefit is, however, assumed to increase in accordance with the change induced, i.e., by one per cent of the initial **user** benefit.
10. ρ^2 is an informal goodness-of-fit measure, defined by

$$\rho^2 = 1 - \frac{\ell(\hat{\eta})}{\ell(\mathbf{O})},$$

where $\ell(\eta)$ is the log-likelihood function, which is maximized for $\hat{\eta} = \eta$.

11. To avoid circularity of argument, whenever a "non-chosen" project is located in the same municipality as the "chosen" project, this variable is corrected so as to take on identical values as between the "chosen" and "non-chosen" project. Without this correction, the "non-chosen" project would in all of these cases receive a higher value on the MUNREP variable than the "chosen" one, simply by construction, translating into a spuriously significant coefficient.

    Note that cases in which there are zero previous representations have been coded so that MUNREP = 1, making the logarithmic transformation allowable. The difference between zero and one representation is captured by the MUNREPZERO dummy.
12. Models including even larger variable sets than the "ample" model were also explored, however without providing significantly improved explanatory power.
13. An accident blackspot is defined as a maximally 100 m long road section in which four or more injury accidents have occurred during 1983–86, or a maximally 1 000 m long road section with ten or more injury accidents.
14. In these models, the linear formulation appears superior to the logarithmic one.
15. The cost-benefit analyses do, in fact, distinguish between these items, business travel time being valued at 3.8 times the (per person) value of commuting time and 6.7 times the value of other (leisure) travel time. However, the resulting composite benefit mesure is not itemized in the final National Road Plan.
16. Odeck's (1996) questionnaire pertains to the 1994–97 planning period, while our data refer to the previous planning period. In principle, this could explain the discrepancy of results. There are reasons to doubt, however, that planning practices have changed substantially between the two planning periods.
17. As for argument (a), however, one may note – again – that this item is, at least in principle, included in the cost-benefit evaluation, and, secondly, that there is very little scientific or empirical evidence supporting the claim that road construction does promote regional economic development in Norway (Lian, 1995).

References

Ben-Akiva, M. and Lerman, S.R. (1985). *Discrete choice analysis: Theory and application to travel demand.* Cambridge, MA: MIT Press.

Chapman, R.G. and Staelin, R. (1982). Exploiting rank ordered choice set data within the stochastic utility model. *Journal of Marketing Research* 19: 288–301.

Elvik, R. (1993). Hvor rasjonell er trafikksikkerhetspolitikken? Report 175. Institute of Transport Economics, Oslo.

Elvik, R. (1995). Explaining the distribution of state funds for national road investments between counties in Norway: Engineering standards or vote trading? *Public Choice* 85: 371–388.

Jansson, J.O. and Nilsson, J.-E. (1989). Spelar samhällsekonomiska kalkyler någon verklig roll i vägväsendet? *Ekonomisk Debatt*, No 2: 85–95.

Lian, J.I. (1995). Næringslivets nytte av infrastrukturinvesteringer. Working Report 998. Institute of Transport Economics, Oslo.

McFadden, D. (1975). The revealed preference of a government bureaucracy: Theory. *Bell Journal of Economics* 6: 401–416.

McFadden, D. (1976). The revealed preference of a government bureaucracy: Empirical evidence. *Bell Journal of Economics* 7: 55–72.

McFadden, D. (1978). Modelling the choice of residential location. In A. Karlquist et al. (Eds.), *Spatial interaction theory and residential location*, 75–96. Amsterdam: North-Holland.

Nilsson, J.-E. (1991). Investment decisions in a public bureaucracy: A case study of Swedish road planning practices. *Journal of Transport Economics and Policy* 10: 163–175.

Nyborg, K. and Spangen, I. (1996). Politiske beslutninger om investeringer i veger. Working Report 1026. Institute of Transport Economics, Oslo.

Odeck, J. (1991). Om nytte-kostnadsanalysenes plass i beslutningsprosessen i vegsektoren. *Sosialøkonomen*, No 3: 10–15.

Odeck, J. (1996). Ranking of regional road investment in Norway: Does socioeconomic analysis matter? *Transportation* 23: 123–140.

Essay 2

Measuring the contribution of randomness, exposure, weather, and daylight to the variation in road accident counts

Pergamon

Accid. Anal. and Prev., Vol. 27, No. 1, pp. 1–20, 1995

0001-4575/95 $9.50 + .00

0001-4575(94)E0023-E

MEASURING THE CONTRIBUTION OF RANDOMNESS, EXPOSURE, WEATHER, AND DAYLIGHT TO THE VARIATION IN ROAD ACCIDENT COUNTS

LASSE FRIDSTRØM,[1] JAN IFVER,[2] SIV INGEBRIGTSEN,[1] RISTO KULMALA,[3] and LARS KROGSGÅRD THOMSEN[4]

[1]Institute of Transport Economics, P.O. Box 6110 Etterstad, N-0602, Norway; [2]Swedish National Road Administration, Röda vägen 1, S-781 87 Borlänge, Sweden; [3]Technical Research Centre of Finland, Sähkömiehentie 3, SF-02150 Espoo, Finland; [4]Danish Council of Road Safety Research, Ermelundsvej 101, DK-2820 Gentofte, Denmark

(*Accepted* 9 *February* 1994)

Abstract—Road accident counts are influenced by random variation as well as by various systematic, causal factors. To study these issues, a four-country, segmented data base has been compiled, each segment consisting of monthly accident counts, along with candidate explanatory factors, in the various counties (provinces) of Denmark, Finland, Norway, or Sweden. Using a generalized Poisson regression model, we are able to decompose the variation in accident counts into parts attributable to randomness, exposure, weather, daylight, or changing reporting routines and speed limits. To this purpose, a set of specialized goodness-of-fit measures have been developed, taking explicit account of the inevitable amount of random variation that would be present in any set of accident counts, no matter how well known the accident generating Poisson process. Pure randomness is seen to "explain" a major part of the variation in smaller accident counts (e.g. fatal accidents per county per month), while exposure is the dominant systematic determinant. The relationship between exposure and injury accidents appears to be almost proportional, while it is less than proportional in the case of fatal accidents or death victims. Together, randomness and exposure account for 80% to 90% of the observable variation in our data sets. A surprisingly large share of the variation in road casualty counts is thus explicable in terms of factors not ordinarily within the realm of traffic safety policy. In view of this observation, it may seem unlikely that very substantial reductions in the accident toll can be achieved without a decrease in the one most important systematic determinant: the traffic volume.

1. INTRODUCTION: THE NEED TO UNDERSTAND ACCIDENT COUNTS

Road accident statistics are, understandably, the subject of considerable interest on the part of media, policy makers, organizations, and the general public. Instances in which accident counts are, for some reason, unusually high receive particular attention. Such cases are almost invariably interpreted as a change in the underlying accident risk and tend to generate some form of public action or outcry.

But accidents counts are influenced by numerous factors other than the risk level. First and foremost, they are subject to random variation. Second, they are strongly influenced by—perhaps almost proportional to—exposure levels. Third, they are affected by natural phenomena like weather and daylight. Fourth, they depend on the accident reporting routines currently in effect and on the changes occurring in these routines over time.

The aim of this research has been to assess how much variation in the accident counts is typically attributable to the above four general factors. Only when all of these factors have been controlled for can we interpret changes in accident counts as attributable to changes in risk, i.e. in the expected (long-term) number of accidents or victims per unit of exposure.

2. A PROBABILISTIC CAUSAL LAW GOVERNING ROAD ACCIDENTS*

Accident counts taken at given points or (smaller) segments of the road system typically ex-

*This discussion draws heavily on the arguments put forward by Fridstrøm (1991, 1992) and Fridstrøm and Ingebrigtsen (1991).

hibit a quite pronounced and apparently random variation from one time period to the next.

Yet, the number of accidents recorded within reasonably large geographical units (e.g. all of Sweden) appears to be almost constant from one year to the next. There is, in general, a striking stability observable in highly aggregate accident counts.

Obviously, there must be a reason why aggregate accident data exhibit such striking stability. Also, there must be a reason why accidents are more frequent in some places or areas than in others. How can the idea of unpredictability and randomness be reconciled with the assertion that accidents have causes, which, if eliminated or weakened, would enable us to reduce the accident toll?

Following the spectacular success of Newtonian mechanics, the notion of causal relations as necessarily deterministic went practically unchallenged for almost two centuries. With the advent of quantum mechanics the situation was, however, radically changed. The behavior of elementary particles is not explained by modern physical theories. Indeed, the predominant view of physical scientists today is that their behavior cannot be explained or predicted. It is not only epistemically, but objectively (ontologically) random in character, i.e. the unpredictability of elementary particles is a feature of the world as it really is, and not only of how we (fail to) understand it. It is not merely a reflection of our incomplete knowledge or measurement technology. No matter how far science proceeds, we would not—according to this tenet—be able to predict events at the micro (particle) level.

However, the collection of elementary particles behaves in a perfectly predictable and stable way, so that the matter made up by the thousands and millions of particles exhibits, in fact, known properties. That is, the properties are known up to a random "disturbance" term, with a given probability distribution. For instance, a radioactive isotope decays, by the emission of neutrons, towards a stable form. It is impossible to say when a specified atom of the isotope will decay, but if we observe a large collection of atoms, we can predict, with astonishing (and known) precision, how long it will take before half of them have decayed. Archaeologists take advantage of this fact to date their finds by means of the so called C_{14} method.

Quantum mechanics therefore introduces an unavoidable element of unpredictability and randomness in science. This has had profound implications for the way in which we view the world as well as for the ways in which we attempt to learn about it. It has become generally accepted among physicists and philosophers of science that causal explanations need not have the form of deterministic relationships in order to be considered complete (Suppes 1970; Papineau 1978, 1985; Salmon 1984; Cartwright 1989). Probabilistic laws of causation, leaving a certain amount of variation "unexplained", are good enough. Certain laws and phenomena are—in the words of Salmon (1984)—"irreducibly statistical" in nature, i.e. they include an objectively (ontologically) random component. In such cases, any attempt to explain more than a certain amount of variation would be futile. If "successful", such explanations would necessarily involve a fallacy, in that certain empirical correlations resulting from sheer coincidence have been misinterpreted as causally determined.

How does this carry over to the social or behavioral sciences? The use of random disturbance terms and probabilistic modeling has become commonplace in economics as well as in other social science disciplines. To the extent, however, that the human being is believed to have a free will, it seems hard to argue that the behavior of the elementary units of society (i.e. its individuals) is random in any other sense than being unknown or unpredictable to the analyst. The collection of individuals may very well behave according to certain social or economic laws, knowable up to a margin of statistical "error". This error is, however, epistemic rather than objective.

In the case of road accidents, on the other hand, there seems to be something more to be said. Although accidents are the result of human choices and behavior, they are not chosen (except for suicidal ones). On the contrary—when an accident happens, it is because certain road users (the accident victims) did not succeed in avoiding it, although they certainly did want to. Accidents are the unintentional side effects of certain actions taken for other reasons than that of causing injury or damage. They are random and unpredictable in the striking sense that had they been anticipated, they would most probably not have happened. Each single accident is, in a sense, unpredictable by definition. We venture the assertion that, nowhere within the realm of behavioral science, is there a set of phenomena coming closer than road accidents to being objectively (ontologically) random in character. No matter how much we learn about accident generating mechanisms or countermeasures, we would never be able to predict exactly where, when, and by whom the single accident is going to occur. Accidents are random in a much more fundamental sense than the conscious choices ordinarily made by social or economic agents. The best we can hope to achieve, is to predict their approximate number.

This suggests that any analysis of the accident generating process should be based on an explicitly probabilistic model—a(n irreducibly) statistical law—according to which single events may occur at random intervals; however, with an almost constant overall frequency in the long run. Although the single event is all but impossible to predict, the collection of such events may very well behave in a perfectly predictable way, amenable to description by means of precise mathematical-statistical relationships. We believe that this principle applies to traffic accidents as it does to quantum physics.

We can, through the design of road systems and vehicles and through our choice of behavior as road users, influence the probability of an accident occurring, thereby altering the long-term accident frequency. This long-term accident frequency—the expected number of accidents per unit of time—we prefer to think of as the result of a causal process. This process accounts for the rather striking stability observable in aggregate accident data, in which the random factors ("noise", "disturbance") having a decisive effect at the micro level, are "evened out" by virtue of the law of large numbers. The causal process determines the expected number of accidents, as a function of all the factors making up the causal set (the causes).

Now, let $\lambda(r, t)$ denote the expected number of accidents occurring during period t in region r. The expected number of accidents is, of course, not a constant—it varies in time and space, i.e. with r and t. We shall refer to this variation, attributable to the various causal factors, as *systematic*. Unlike the random or pure chance variation, the systematic variation can—in principle—be influenced and controlled. Only the systematic variation is of interest from a policy point of view.

Assume that the probability that an accident occurs in area r during a given (short) time interval is constant throughout period t, and independent of any previous accident events. In this case the observed accident number ($y(r,t)$, say) can be shown to follow a Poisson probability law, given by

$$P[y(r,t) = m] = \frac{[\lambda(r,t)]^m \cdot e^{-\lambda(r,t)}}{m!}.$$

Although in practice we will be working with observation periods as long as one month, the assumption of constant accident probability throughout the observation period is, in fact, an innocuous one. This is so on account of the convenient invariance-under-summation property of the Poisson distribution: any sum of independent Poisson variates is itself Poisson distributed, with parameter (λ) equal to the sum of the underlying, individual parameters. Thus all we need to assume is that, for some very short time interval (say, a minute or a second), the accident probability can be considered constant and that events occurring during disjoint time intervals are probabilistically independent. In such a case the number of accidents occurring in a given region r during a given month t will, indeed, be Poisson distributed with some (unknown) parameter $\lambda(r,t)$.

The Poisson is a one-parameter distribution, with the very interesting property—crucial to our analysis—that the variance equals the expected value, both being equal to the Poisson parameter $\lambda(r,t)$. That is, knowing the expected value, one also knows, in a sense, how much random variation is to be expected around that expected value. In fact, one knows the entire distribution.

Now, to identify and estimate the effects of systematic factors on the accident counts, we specify

$$\lambda(r,t) = e^{\Sigma_j x_j(r,t)\,\beta_j}$$

i.e., $\ln[\lambda(r,t)]$ can be written as a linear regression determined by a set of independent variables x_j and a set of coefficients $\beta_j (j = 1,2, \ldots ,J)$. This choice of functional form is in a sense a natural one, in that it makes sure that the expected number of accidents is always a positive number, although possibly a very small one. To the extent that the x_j variables are measured on a logarithmic scale, the β_js are interpretable as (constant) "accident elasticities", i.e. as the percentage increase in the expected number of accidents $\lambda(r,t)$, following a 1% increase in x_j. For x_j variables measured on an ordinary linear scale, the elasticity is given by $x_j \cdot \beta_j$, i.e. increasing in x_j. For dummy x_j variables, the β_js approximately measure the relative increase in $\lambda(r,t)$ as the dummy variable changes from zero to one.

Now assume, for the sake of the argument, that we have somehow acquired complete and correct knowledge of all the factors x_j causing systematic variation, and of all their coefficients β_j. In other words, the expected number of accidents $\lambda(r,t)$—i.e. all there is to know about the accident generating process—is known. Could we then predict the accident number with certainty? The answer is no: there would still be an unavoidable amount of purely random variation left, the variance of which would be given—precisely—by $\lambda(r,t)$. The residual variation should never be smaller than this, or else one must conclude that part of the purely random variation has been misinterpreted as systematic and erroneously attributed to one or more causal factors.

Now, in practice one is seldom in the fortunate situation that all risk factors have been correctly identified and their coefficients impeccably estimated, so that the expected number of accidents is virtually known. A somewhat more realistic probability model arises if one assumes that the Poisson parameter $\lambda(r,t)$ is itself random, and drawn from a gamma distribution with shape parameter ξ (say), in which case the observed number of accidents can be shown (Gourieroux, Montfort, and Trognon 1984 a, b) to follow a negative binomial distribution with expected value $E[\lambda(r,t)] = \mu(r,t)$ (say) and variance

$$\sigma^2(r,t) = \mu(r,t)\,[1 + \theta\,\mu(r,t)],$$

where $\theta = 1/\xi$.

In the special case $\theta = 0$, the gamma distribution is degenerate, and we are back to the simple Poisson distribution, in which the variance equals the mean. We shall refer to θ as the *overdispersion parameter*, and to models in which $\theta > 0$ as *overdispersed.* In such a model, the amount of unexplained variation is larger than the normal amount of random disturbance in a perfectly specified Poisson model, meaning, in fact, that not all the noise is purely random. The model does not explain all the systematic variation, but lumps part of it together with the random disturbance term.

The overdispersion parameter can be used to test whether or not our independent variables explain all the explicable (systematic) variation, i.e. if there is residual variation left in the model over and above the amount that should be there in a perfectly specified and estimated Poisson model. However, even if the overdispersion is found to be zero, it does not follow that the analyst has found all the true causal factors and correctly calculated their effects. Our formulation is no guarantee against spurious correlation being interpreted as causal, only against too much correlation being interpreted that way. In principle, two quite distinct sets of alleged causal factors could provide equally good and apparently complete explanations, as judged by the overdispersion criterion. As in other econometric work, the choice of independent variables must be guided by theory and professional judgement, rather than by curve-fitting.

3. A FOUR-COUNTRY DATA SET

To estimate these generalized Poisson regression models, combined cross-section/time-series data bases have been compiled for each of the four greater Nordic countries. Monthly accident counts are given for each county (province), of which there are 14 in Denmark, 11 in Finland, 19 in Norway, and 24 in Sweden. The time periods of observation used for this study extend from 1977 through 1987 in the case of Denmark, from 1975 through 1987 for Finland, from 1973 through 1986 for Norway, and from 1976 through 1987 for Sweden, yielding, respectively, 1848, 1716, 3192, and 3456 units of observation.

Apart from accident statistics, the data bases include data on gasoline sales (a proxy for exposure or traffic volume), weather conditions, the duration of daylight, changes in legislation and reporting routines (dummies), a trend variable, dummy variables for the different counties and months, as well as a number of data items that have not been utilized in the present analysis. Due to dissimilarities with respect to the availability and quality of statistical sources in the four countries, it has not been possible to adopt exactly the same variable definitions and classifications in all countries, nor has it been possible to lump all data into one four-country data base. Thus, we have not been able to analyze the variation between countries; only the temporal and spatial variation within each country has been subject to study.

In Denmark, data on monthly fuel sales are not available for each county. Instead, traffic counts pertaining to certain cross-sections of the road system are used as measures of temporal variation in exposure, while the spatial variation (between counties) in exposure has been estimated on the basis of regional road use statistics for the year 1980.

Meteorological data are available with a differing degree of detail—in some countries only monthly averages (on temperature and precipitation) exist, while in others we have been able to record, e.g. the number of (half-)days (i.e. 12- or 24-hour periods) during which the temperature drops below the point of freezing, and/or the number of days with precipitation in the form of rain or snow. These data are based on the records collected by the many meteorological stations in operation, of which there are usually several in each county. A selection of stations has therefore had to be made, based on the completeness of available records and the proximity to the county's "center of gravity", as measured in terms of traffic volumes. Most counties are, however, small enough that the weather records would be only marginally different between different stations. In some cases, however, the average of several meteorological stations within the county is used, to account for weather variations within geographically extended and heterogeneous regions.

The amount of daylight per 24-hour period has been compiled on the basis of almanac data on the exact times of sunrise and sunset, as measured at a sample point in each county. In the Nordic countries, this variable exhibits an unusual amount of variation, cross-sectionally as well as over the year. In the northernmost regions, e.g. the measured amount of daylight is zero in December–January and 24 hours a day in June–July.

For Finland, daylight is measured from midnight till midnight. For Denmark, Norway and Sweden, however, the daylight data pertain to the period between 7 A.M. and 11 P.M. In these three countries, therefore, the daylight measures are affected by the introduction, from 1980 onwards, of daylight savings time during April through September and by the fact that sunrise and sunset occur at an earlier clock time in the easternmost counties. The latter effect is particularly pronounced in Norway, where the solar time difference between the easternmost and westernmost points is almost two hours; yet the entire country is in one time zone.

Models have been estimated with three types of dependent variables: the number of injury accidents (I), the number of fatal accidents (F), and the number of road users killed (K). An accident is called fatal if one or more persons are killed. Injury accidents are accidents in which one or more persons are killed *or* injured, i.e. the I category includes even the fatal accidents.

For Denmark data are lacking on the number of road users killed, while for Finland data were not available on fatal accidents.

4. GENERALIZED POISSON REGRESSION MODELS FOR FOUR COUNTRIES

4.1. *Overview of models estimated*

A large number of statistical models have been estimated, differing with respect to the set of dependent and independent variables used. An overview of models presented in this paper is given in Table 1, in which we also exhibit the notation used to refer to the various models.

The simplest models (I1, F1, K1) are the ones containing only a constant term and a measure of exposure (usually gasoline sales). In a second set of models (I2, F2, K2) we include dummy variables capturing important changes in reporting routines affecting the accident statistics, and in traffic legislation. In a third step, variables describing the weather conditions and the amount of daylight prevailing in a particular county and month are added. Fourth, we add a linear trend factor, capturing gradual changes in the risk level, uniformly for all counties in a given country. Fifth, we add a set of regional dummy variables, one for each county except one. In the most complete set of models ($I6$, $F6$, $K6$), a similar set of seasonal dummies, one for each month except December, is added.

These two sets of dummy variables are meant to represent whatever regional and seasonal differences in the accident risk level are not captured by the other variables included. They are, however, not uncorrelated with these variables, and tend to absorb a large part of the variation attributable to weather, daylight, and—in particular—exposure, possibly distorting those coefficients heavily. For purposes of causal inference we therefore prefer models $I4$, $F4$, and $K4$, as discussed in section 4.2 below.

Apart from the linear trend term and certain changes in the Danish speed limit legislation, no attempt has been made, in this study, to estimate the effect of risk or safety factors other than weather and daylight. This is so, although certain variables of interest are, in fact, included in the data base and could have been put into the models. We believe, however, that in order to estimate reliably the partial effects of such factors, a fairly complete set of factors would have to be included (in principle, all the factors influencing risk), or else the coefficients of the variables included would be subject to major omitted-variable bias, since the social, demographic, economic, and policy variables of interest would be highly correlated in a data set like ours. The weather and daylight variables, on the other hand, are necessarily exogenous to the economic, social, and accident-generating process, and hence should not give rise to any important bias of this kind. If any correlation exists between these variables and any set of omitted "explanatory" factors, it must be because the omitted variables are influenced by weather and daylight–or at least by their normal seasonal and geographic pattern of variation (the climate)—rather than vice versa. This amounts to saying that, at worst, the weather and daylight coefficients incorporate, not only the immediate effect on casualty risk of e.g. rainfall and snowfall, but also—to some extent—the effects of certain economic and social variables having, on account of climatic factors, a clear seasonal or geographic pattern of variation.

The models were estimated by means of the LIMDEP 5.1 computer software (Greene 1990), using the maximum likelihood estimation method.

4.2. *Estimated systematic effects*

In Tables 2 through 4 we present the coefficients of these models, explaining injury accidents, fatal accidents, and road users killed, respectively, as a

Table 1. Overview of models estimated

Independent variables	**Model notation (* means variable(s) included)**						
Months (dummies)							*
Counties (dummies)						*	*
Linear trend					*	*	*
Weather and daylight				*	*	*	*
Reporting and legislation			(*)	(*)	(*)	(*)	(*)
Exposure		*	*	*	*	*	*
Constant	*	*	*	*	*	*	*
Country	**I - models explaining injury accidents**						
Denmark	DK-I0	DK-I1	DK-I2	DK-I3	DK-I4	DK-I5	DK-I6
Finland	SF-I0	SF-I1	SF-I2	SF-I3	SF-I4	SF-I5	SF-I6
Norway	N-I0	N-I1	N-I2	N-I3	N-I4	N-I5	N-I6
Sweden	S-I0	S-I1	S-I2	S-I3	S-I4	S-I5	S-I6
	F - models explaining fatal accidents						
Denmark	DK-F0	DK-F1	DK-F2	DK-F3	DK-F4	DK-F5	DK-F6
Finland							
Norway	N-F0	N-F1		N-F3	N-F4	N-F5	N-F6
Sweden	S-F0	S-F1		S-F3	S-F4	S-F5	S-F6
	K - models explaining road users killed						
Denmark							
Finland	SF-K0	SF-K1		SF-K3	SF-K4	SF-K5	SF-K6
Norway	N-K0	N-K1		N-K3	N-K4	N-K5	N-K6
Sweden	S-K0	S-K1		S-K3	S-K4	S-K5	S-K6

function of exposure, reporting/legislation, weather, daylight, and trend.

Exposure. For injury accidents, exposure, as proxied by the gasoline sales, comes out with a coefficient close to one, meaning near proportionality between exposure and the expected number of injury accidents. For Denmark, two separate coefficients are estimated, one capturing time-series and the other cross-sectional variation. Here, only the cross-section effect is seen to be close to one. For the other three countries, the exposure variable captures cross-section as well as time-series variation.

For fatal accidents and death victims (Tables 3 and 4), the exposure coefficients are significantly smaller than one, suggesting a less than proportional relationship between traffic volume and casualties. In Sweden, for examples the expected number of fatal accidents increases by an estimated 0.64% for each percent increase in the gasoline sales. In other words, the risk per exposure unit decreases by about 0.36% when the traffic volume grows by 1%.

There are several ways to interpret this result. One interesting hypothesis is that the average severity of accidents decreases with the traffic volume, since the average speed level is forced down in denser traffic. This interpretation seems consistent with the findings of Fridstrøm and Ingebrigtsen (1991), who find that increases in traffic density (as measured in terms of fuel sales per unit road length), for given exposure, tends to reduce the number of casualties, especially the fatal ones. Also, such an hypotheses might help understand why the time-series effect (as estimated on the Danish data set) is considerably smaller than the cross-section effect. Over time, the length of the road network in a given county is almost constant, meaning that the effect of increased exposure is dampened by the (opposite) effect of increased density. In the cross-section of counties, the association between traffic volume and traffic density is much less pronounced.

Another possible interpretation of the less than proportional relationship between exposure and casualties, is by reference to learning (Adams 1987). As society becomes increasingly more used to motorized transport, knowledge on how to avoid accidents, reduce their severity, or repair the damage caused accumulates among individuals and institutions. Driving proficiency tends to improve, the road system is amended, vehicle crashworthiness is enhanced, and medical advances and improvements in the public health system account for increased survival rates among accident victims with very severe injuries.

The average fuel efficiency of the car park has

Table 2. Models I4: injury accidents (coefficient estimates, with standard errors in parentheses)

j Variable	Denmark	Finland	Norway	Sweden
0 Constant	−1.301(.149)	1.665(.052)	−3.991(.101)	1.956(.045)
Exposure				
1 ln(temporal traffic index by county)	0.394(.043)			
2 ln(traffic level in county as of 1980)	0.979(.012)			
3 ln(gasoline sales*)		1.051(.010)	0.946(.010)	0.990(.008)
Reporting and legislation (dummies)				
4 1.3.1979: new speed limits	−0.190(.021)			
5 1.10.1985: new speed limits	0.003(.021)			
6 1.1.1978: new reporting routines		−0.234(.020)		
7 1.1.1977: new accident report forms			0.033(.017)	
8 1.10.1978: new reporting routines			−0.137(.019)	
9 1.1.1985: new accident report forms				−0.039(.016)
Weather and daylight				
10 monthly precipitation (mm/1000)	0.881(.178)			
11 monthly days with precipitation/100				0.239(.120)
12 monthly days with rainfall/100		1.156(.146)	0.274(.112)	
13 monthly days with snowfall/100	−1.219(.150)	−0.479(.152)	−0.048(.157)	
14 sudden snowfall (dummy)	0.050(.016)	0.015(.024)	0.042(.034)	
15 monthly days with frost/100			−1.753(.124)	−1.554(.081)
16 monthly half-days with frost/200	−0.933(.137)			
17 snow depth (cm/100)		−0.472(.039)	0.035(.033)	−0.410(.045)
18 minutes of daylight per day/1000		−0.362(.035)		
19 minutes of daylight 7 am–11 pm/1000	−0.105(.053)		−0.710(.038)	−0.655(.038)
Trend				
20 months since first observation/100	−0.174(.030)	−0.029(.019)	−0.160(.022)	−0.091(.018)
Summary sample statistics				
Overdispersion parameter	0.036(.002)	0.032(.002)	0.044(.002)	0.052(.002)
Log-likelihood	−7376.2	−6497.4	−11558	−13538
Sample size	1848	1716	3192	3456
Degrees of freedom	1836	1706	3180	3447
Mean of dependent variable	64.0	56.8	37.5	55.1
Maximum of dependent variable	193	234	163	314
Minimum of dependent variable	2	7	0	0
Maximum of predicted values	164.6	222.0	125.7	349.5
Minimum of predicted values	4.059	10.5	6.120	4.677

*Measured in thousands of liters for Norway, and millions of liters for Finland and Sweden.

improved over time, although in the Nordic countries at a rather moderate pace. In Norway, for example, gasoline consumption per vehicle kilometer went down by an estimated 13% between 1975 and 1987 (Rideng 1993). Thus the increase in exposure over time is slightly understated in our data sets.

The learning and traffic density effects would tend to translate into a less than proportional relationship between gasoline sales and casualties, as estimable in a time-series data set. The fuel efficiency effect, on the other hand, would pull the exposure coefficient in the opposite direction. Now, in our data sets the amount of time-series variation is small compared to the cross-sectional variation. Moreover, the influence of any factor with a spatially stable, but temporally monotonic pattern of variation is likely to be captured, to a rather large degree, by the linear trend term (see below). The biases due to learning and fuel efficiency effects are, therefore, probably not very large.

It might be argued that, in models like these, the exposure coefficient ought to be constrained to one. In such a case, the remaining parameters of the model would be interpretable as estimates of the pure effects on risk, rather than as a mixture of effects on risk as well as exposure. This formulation, however, would be tantamount to assuming strict proportionally between exposure and accidents, something that appears unreasonably restrictive in view of the above arguments and notably in view of the empirical results derived here.

Accident reporting. Accident reporting routines may have an artificial, though significant, effect on casualty counts. In each of the four countries, all road accidents involving injury are, in principle, subject to mandatory police reporting. Road accident statistics are compiled on the basis of these police reports. It is well known, however, that the reporting is far from complete, except probably for fatal accidents/death victims (Finnish Roads and Waterway Administration 1982; Borger 1991; Nedland and Lie 1986; Thulin 1987; Gothenburg City Planning Office 1986; Larsen 1989). Certain categories of accidents are reported less frequently than

Table 3. Models F4: fatal accidents (coefficient estimates, with standard errors in parentheses)

j Variable	Denmark	Finland	Norway	Sweden
0 Constant	−2.640(.433)		−3.532(.300)	0.123(.101)
Exposure				
1 ln(temporal traffic index by county)	0.432(.116)			
2 ln(traffic level) (1980)	0.802(.037)			
3 ln(gasoline sales*)			0.575(.030)	0.640(.018)
Reporting and legislation (dummies)				
4 1.3.1979: new speed limits	−0.234(.052)			
5 1.10.1985: new speed limits	−0.050(.054)			
Weather and daylight				
10 monthly precipitation (mm/1000)	0.984(.431)			
11 monthly days with precipitation/100				0.317(.276)
12 monthly days with rainfall/100			−0.384(.328)	
13 monthly days with snowfall/100	−1.467(.394)		−0.989(.467)	
14 sudden snowfall (dummy)	0.049(.038)		0.062(.086)	
15 monthly days with frost/100			−1.029(.363)	−1.313(.184)
16 monthly half-days with frost/200	−0.762(.365)			
17 snow depth (cm/100)			−0.296(.109)	−0.519(.120)
19 minutes of daylight 7 am–11 pm/1000	−0.530(.129)		−0.742(.116)	−0.712(.085)
Trend				
20 months since first observation/100	0.066(.076)		−0.328(.030)	−0.303(.025)
Summary sample statistics				
Overdispersion parameter	0.066(.011)		0.030(.014)	0.030(.009)
Log-likelihood	−3804.1		−5115.1	−6358.1
Sample size	1848		3192	3456
Degrees of freedom	1836		3182	3448
Mean of dependent variable	3.685		1.740	2.727
Maximum of dependent variable	17		11	16
Minimum of dependent variable	0		0	0
Maximum of predicted values	8.435		4.196	9.857
Minimum of predicted values	0.399		0.467	0.463

*Measured in thousands of liters for Norway, and millions of liters for Sweden.

Table 4. Models K4: road users killed (coefficient estimates, with standard errors in parentheses)

j Variable	Denmark	Finland	Norway	Sweden
0 Constant		0.043(.123)	−3.324(.300)	0.146(.109)
Exposure				
3 ln(gasoline sales*)		0.778(.023)	0.557(.031)	0.641(.021)
Weather and daylight				
11 monthly days with precipitation/100				0.553(.291)
12 monthly days with rainfall/100		1.014(.334)	−0.458(.336)	
13 monthly days with snowfall/100		0.667(.353)	−0.860(.483)	
14 sudden snowfall (dummy)		−0.032(.048)	0.044(.100)	
15 monthly days with frost/100			−1.080(.374)	−1.188(.196)
17 snow depth (cm/100)		−0.864(.090)	−0.251(.112)	−0.447(.121)
18 minutes of daylight per day/1000		−0.393(.076)		
19 minutes of daylight 7am–11pm/1000			−0.689(.120)	−0.634(.092)
Trend				
20 months since first observation/100		−0.371(.028)	−0.299(.032)	−0.304(.027)
Summary sample statistics				
Overdispersion parameter		0.060(.009)	0.157(.018)	0.123(.010)
Log-likelihood		−3797.4	−5507.2	−6897.7
Sample size		1716	3192	3456
Degrees of freedom		1707	3182	3448
Mean of dependent variable		4.776	1.915	3.058
Maximum of dependent variable		28	12	18
Minimum of dependent variable		0	0	0
Maximum of predicted values		17.420	4.422	10.780
Minimum of predicted values		0.987	0.559	0.548

*Measured in thousands of liters for Norway, and millions of liters for Finland and Sweden.

others. It cannot be excluded, therefore, that changes in the incidence of reporting correlate with changes in one or more of our independent variables, thus biasing the coefficients of the latter. Accidents involving bicyclists, for example, are known to be subject to major underreporting, thus biasing the coefficients of any variable correlated with bicyclist exposure (such as weather and daylight).

Certain known, important changes in the reporting routines have been incorporated in the models in the form of dummy variables. In Finland, for example, the introduction of new routines in January 1978 seems to have reduced the accident counts by an estimated 21% ($1 - e^{-.234}$). In Norway, accidents with only "minor" injury have not been subject to mandatory reporting since October 1978; this appears to have reduced the injury accident count by some 13%.

Legislation. During the period of observation considered, important legislative changes have taken place only in Denmark, in the form of lowered speed limits. As of March 1st, 1979, the speed limits on rural roads and freeways were lowered from 90 to 80 km/h and from 110 to 100 km/h, respectively. From October 1st, 1985, the urban speed limit in Denmark was lowered from 60 to 50 km/h. The first speed limit reduction appears to have had a significant effect on road casualties, reducing the rate of fatal accidents by an estimated at 21%, while the effect of the second one is statistically insignificant according to our analysis.

Weather. Weather conditions have a significant impact on accident counts, although in some cases the direction of impact may seem counterintuitive. True, rainfall is liable to increase the accident toll. Snowfall, however, seems to have the opposite effect. In Denmark, for example, the expected monthly number of injury accidents decreases by an estimated 1.2% for each additional day of snowfall during the month. An even larger effect is estimated for fatal accidents. Even the incidence of frost comes out with a significantly negative coefficient. Here, however, the effect on fatal accidents/death victims appears to be smaller than on injury accidents in general. (For Sweden data are lacking on the frequency of snowfall, and for Finland on the frequency of frost.)

Several interpretations are possible. Most car drivers in the Nordic countries are well used to traveling under typical winter conditions, and may, under the risk compensation hypothesis, be thought to adjust their driving habits so as to more or less offset the increased hazard due to slippery road surfaces. Indeed, it cannot be ruled out that this behavioral adjustment is more than large enough to keep the risk level constant, whereby a net decrease in the accident count would be observable.

Other, more conventional, lines of explanation include the possible effects of reduced exposure during winter. This effect is not completely controlled for through our gasoline sales variable, (*i*) because the fuel consumption per vehicle kilometer increases when the temperature drops, the road is covered by snow or ice, and/or winter tires are used; and (*ii*) because the exposure due to pedestrians, bicyclists, motorcyclists, or diesel-driven vehicles is not reflected in the gasoline sales statistics.

The plausibility of this explanation is, however, weakened by the fact that the same kind of effects is found even for Denmark, in which exposure is measured directly, by means of traffic counts, rather than indirectly, through the fuel sales statistics.

A third possible interpretation could be that visibility at night is increased when the road(side) is covered by snow. This helps offset the unfavorable safety effect of a slippery surface.

Fourth, reporting bias could, in principle, be the source of almost any statistical correlation found. Note, however, that the bias due to underreporting of bicycle accidents would especially tend to deflate the casualty counts during summer. It cannot, therefore, explain the apparent casualty-reducing effect of wintery weather. Even the effects of daylight and exposure are probably underestimated on account of this.

Fifth, it is conceivable that less proficient drivers tend to refrain from driving during difficult conditions, so that the driving population has a higher average level of proficiency during winter.

Sixth, snow drifts along the roadside may have the effect of dampening the impact of single vehicle crashes (Brorsson, Ifver, and Rydgren 1988), or perhaps prevent the vehicle from leaving the road, so that fewer accidents end up causing injury. Since only injury accidents are covered by our accident statistics, this might have the effect of reducing the accident count. To capture this effect, we have entered snow depth as a separate independent variable for those countries in which such data have been available. The snow depth is seen to have a significantly negative (i.e. favorable) effect on injury accidents in Finland and Sweden, but an insignificant effect in Norway. The effect on fatal accidents or death victims is, however, statistically significant in all three countries. (For Denmark the relevant data are lacking.)

The first snowfall occurring during the winter season may, however, seem to catch the drivers by sufficient surprise to cause an increased accident risk, as witnessed by the generally positive (although hardly significant) coefficients of the "sudden snowfall" variable. We measure this by means of a dummy variable that is set equal to one whenever a snowfall occurs during the current month, but not during the preceding one. Previous studies for Norway have shown more significant effects at this point (Fridstrøm and Ingebrigtsen 1991).

Daylight. The amount of daylight has a quite unambiguous and favorable effect on the expected number of accidents. In the Nordic countries, this factor exhibits an unusual amount of variability, between counties as well as over the year. An extra hour of daylight between 7 A.M. and 11 P.M. corresponds to an estimated 4% decrease in the expected number of injury accidents in Norway ($1 - e^{-0.710 \cdot 60/1000} = 0.04$).

Trend. The linear trend factor comes out with a significantly negative coefficient in eight out of ten cases, the remaining two being statistically insignificant. For injury accidents, the trend factor is estimated at −0.17% per month in Denmark, −0.03% in Finland, −0.16% in Norway, and −0.09% in Sweden, translating into annual risk reduction rates of −2.1%, −0.35%, −1.9%, and −1.1%, respectively. In terms of death victims, the trend effects are larger, and estimated at −4.4%, −3.5%, and −3.6% annually for Finland, Norway, and Sweden, respectively. We interpret the trend variable as a proxy for all those factors that combine to reduce gradually the risk of road casualties over time, be it improvements in the road infrastructure, in the inner safety of cars, in the proficiency and behavior of drivers, or in any other area related to the safety of road users. The gradual improvement in fuel efficiency is also likely to have influenced the trend term, although in the opposite direction.

Constant. In these models, the constant term does not have any interpretation of its own, as it merely reflects the size of the unit of observation (average number of events per county per month) and the unit of measurement used for the independent variables, in particular the fuel sales variable.* To the extent that dummy explanatory variables are included in the model, the constant term will also be affected by the choice of reference category for the dummy variables (i.e., by which category receives the code zero on all dummies).

Overdispersion. The overdispersion parameter is significantly larger than zero in all the models I4/F4/K4, meaning that the independent variables fail to explain all the systematic variation in casualty counts. The overdispersion parameter is, however, generally not very large (except in models K4), meaning that the amount of unexplained systematic variation is relatively small. We revert to this issue in section 5 below.

Regional and seasonal dummies. To account for all those factors that vary across regions and/or according to the season, other than exposure, weather, and daylight, one might consider including into the model a full set of dummy variables for counties and months, i.e. one for each county, except one, and one for each month, except one. When estimated, these models, corresponding to codes I6/F6/K6 in Table 1, are seen to yield coefficient estimates for exposure, weather, and daylight that are generally much smaller (and less significant) than in models I4/F4/K4, suggesting that the causal effects of these variables are, to a large extent, channelled through the regional and seasonal dummies. To check this, another set of models were estimated in which the exposure coefficient was constrained to one (meaning that the models explain the casualty risk per unit of exposure rather than the crude number of casualties). In these models, the regional and seasonal dummies come out substantially smaller than in the otherwise identical models with unconstrained exposure coefficients. In other words, when we constrain the models so as to impose proportionality between casualities and exposure, a large part of the "seasonal" and "regional" variation captured by the dummies disappears.

In general, models I6/F6/K6 exhibit remarkably small overdispersion, some of them no (statistically significant) overdispersion at all, meaning that the models "explain" as much systematic variation as there is to explain.

A note on autocorrelation. Our maximum likelihood method of estimation does not take account of autocorrelation, a potentially serious source of inefficiency in a combined cross-section/time-series data set like ours. While the coefficient estimates themselves remain statistically consistent, their estimated standard errors are probably on the low side, yielding somewhat exaggerated t statistics. Autoregressive count data models have been formulated by Brännäs and Johansson (1992), who, notably, try out their methods on a subset of our data set (the Västerbotten county of Sweden). While these tests do show statistically significant autocorrelation, the error of estimation occurring in models not taking account of autocorrelation appears to be rather

*Fuel sales are measured in thousands of liters for Norway, and millions of liters for Finland and Sweden. This means, for example, that the Norwegian constant term is $\beta_3 \times \ln(1000) = \beta_3 \times 6.908$ smaller than it would have been, had we used the same scale in the Norwegian data set as in the Finnish and Swedish ones. (β_3 is the gasoline sales coefficient.)

small, especially for the negative binomial model corresponding most closely to our formulation.

5. THE "EXPLANATORY POWER" OF DIFFERENT FACTORS

The Poisson regression model, when appropriate, offers some rather interesting opportunities for interpretation, compared to models based on other distributional assumptions. In a perfectly specified and estimated Poisson regression model, i.e. one in which we have actually found the true value of the Poisson parameter determining the distribution of the dependent variable (e.g., the accident intensity) at each sample point, we also know how much deviation from those true, expected values is to be expected due to pure randomness. Having "explained" all the systematic variation, we have, in other words, a way to predict even the random one. The variance of a Poisson variable is equal to its mean. Unlike the situation in most other regression models, it is—in principle—possible to compute an optimal fit, determined by the amount of (purely random) variation that would be present in the perfectly specified model. Any model providing a higher "explanatory power" than this would have to be discarded as "underdispersed" ("overfitted"), meaning that part of the purely random variation has in fact, through our estimation procedure, been treated and interpreted as systematic.

To apply this principle, one has to define some measure of "explanatory power" or "goodness-of-fit". In this paper we define five—rather different—such measures, one based on the log-likelihood ratio, a second based on the overdispersion parameter (the "Elvik index"), a third based on the traditional multiple correlation coefficient (R^2), a fourth (R^2_w) based on the weighted (variance stabilizing) residuals, and a fifth (R^2_{FT}) based on the Freeman-Tukey transformation residuals (see the appendix for a detailed derivation and description). Any choice of goodness-of-fit measure is to some extent arbitrary. By relying on more than one measure we attempt, however, to minimize the risk of drawing conclusions on account of methodological choices rather than subject matter relationships. The results of these calculations are summarized in Figs. 1 through 3.

The part attributable to randomness has been calculated as the amount of sample variation to be expected on account of normal Poisson disturbance only, given that the true Poisson parameters are as estimated under models I4, F4, or K4. This part is very robust with respect to changes in the model specification, since only the sample mean of predicted values is involved in the calculation. The part due to exposure is defined as the fit obtained in models I1/F1/K1, while the part attributable to reporting and legislation is calculated as the additional fit obtained when moving from I1/F1/K1 to I2/F2/K2 (cf. Table 1), etc. In other words, variables are added to the models in a stepwise fashion, in an order given by the upper part of Table 1 (or by Figs. 1–3), as read from the bottom category up. This must be born in mind when the diagrams are interpreted, as the (added) "explanatory power" due to a given varible is strongly dependent on what other independent variables are already present in the model.

The five different ways of measuring "explanatory power" are, by and large, seen to yield reassuringly similar results.

5.1. Models for injury accidents

For injury accidents (Fig. 1), less than 10% of the sample variation can be ascribed to random factors alone (around 3% for Sweden and 9% for Norway). Exposure, on the other hand, is seen to "explain" 65% to 85% of the total variation, and no less than 72% of the "explicable", systematic variation that is left after the purely random part has been subtracted. Reporting and legislation, as measured in our models, account for about 7% in Denmark, less in the other three countries. Weather and daylight are most important in Norway (about 6%). Taken together, these four general factors (randomness, exposure, reporting/legislation, and weather/daylight) are able to "explain" (by the log-likelihood measure) 87% of the variation in Denmark, 94% in Finland, 85% in Norway, and 90% in Sweden. Note that, apart from the speed limit dummies included in the Danish model, there is not a single road safety measure among the list of "explanatory" factors so far included in the models (unless one chooses to regard exposure as a variable open to public intervention).

The trend term, which—in principle—incorporates long-term, linear safety improvements, has a barely noticeable explanatory power. Regional factors as captured by a set of dummy variables (one less than the number of counties) "explain" an additional 8% cent of the variation in Denmark, but less than 1% in Finland. The 11 seasonal dummies add a final 1% to 2% to the goodness-of-fit in each country. Most of the regional and seasonal variation has, of course, already been captured by the exposure, weather, and daylight variables.

5.2. Models for fatal accidents

Turning to fatal accidents (Fig. 2), one finds that the picture is interestingly different. As such

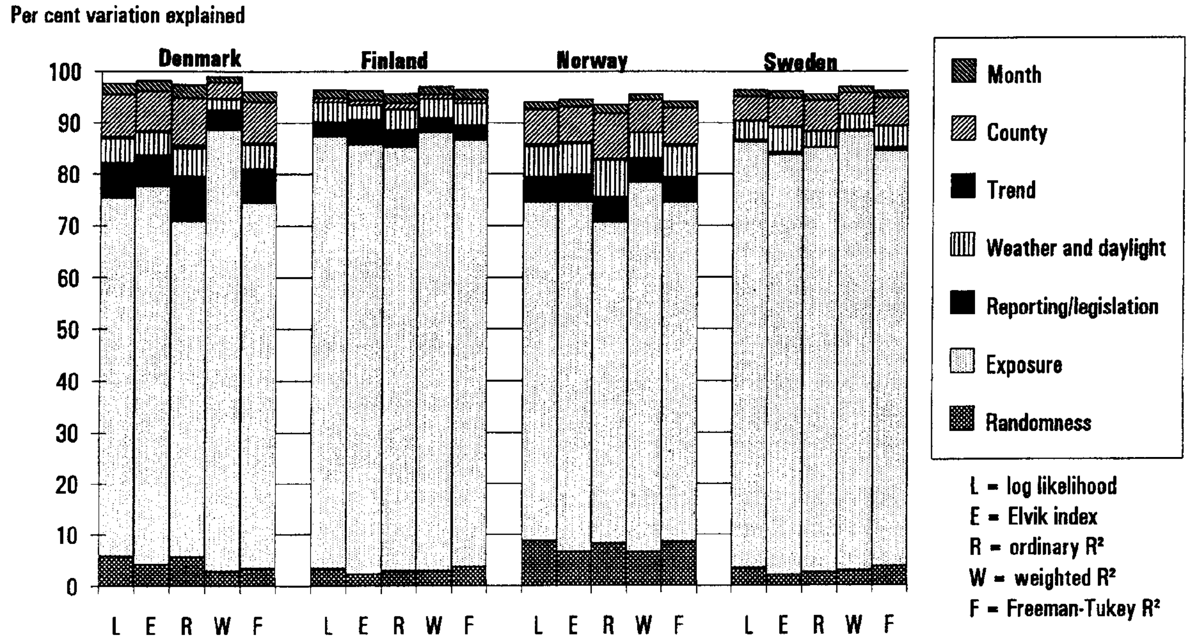

Fig. 1. Variation in injury accidents, decomposed by source, according to five different goodness-of-fit measures.

accidents are a lot fewer, the (relative) scope for random variation is dramatically larger, amounting to no less than 80% in the Norwegian data set and around 60% in the Danish and Swedish data sets. Exposure accounts for about 62% of the remaining (systematic) variation in Denmark, 52% in Norway, and 73% in Sweden. Taken together, randomness, exposure, weather, and daylight explain 87% of the variation in Denmark, a full 94% in Norway, and 93% in Sweden.

For fatal accidents, the trend factor has a non-negligible role to play, at least in Norway and Swe-

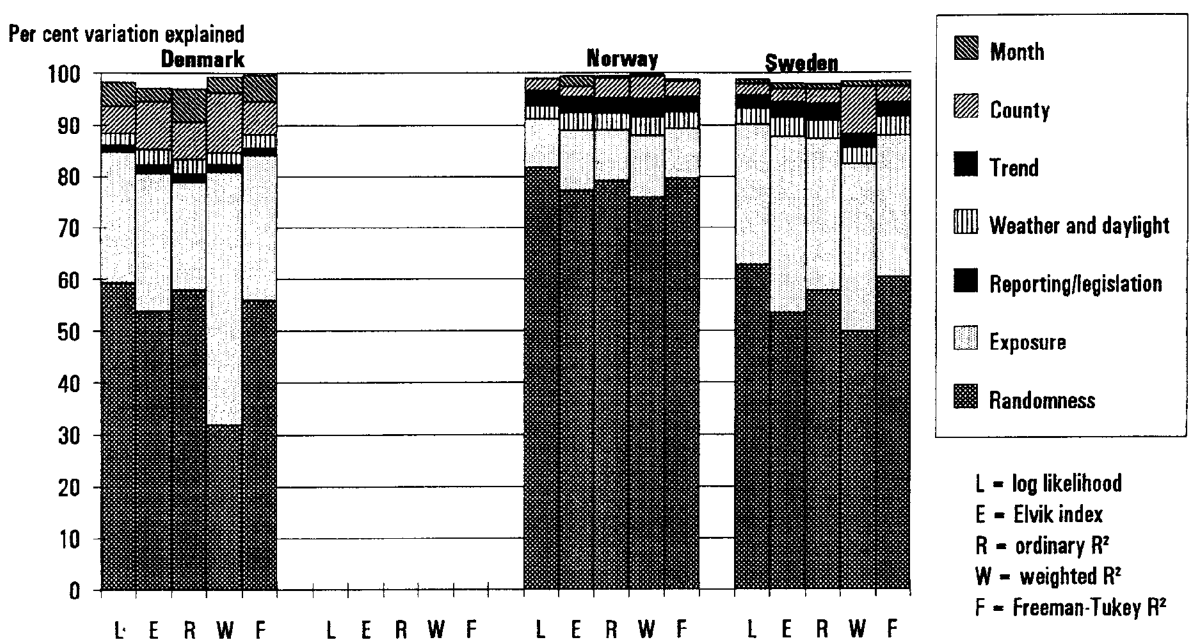

Fig. 2. Variation in fatal accidents, decomposed by source, according to five different goodness-of-fit measures.

den, where its contribution amounts to between 2% and 3%. By assumption, reporting routines do not influence the statistics on fatal accidents. The Danish speed limit measures account, however, for about 1% of the variation in that country.

When the regional and seasonal dummies are added to the model, the amount of "explained" variation becomes remarkably close to 100%. The overdispersion parameter is, in fact, not statistically significant in any of the models DK-F6, N-F6, or S-F6, meaning that the pure Poisson model is fully satisfactory, and there is virtually nothing left to explain beyond the normal random noise. We choose, however, not to place confidence in these models, as, e.g., the exposure coefficients tend to drop to implausibly low levels, and a number of other coefficients become statistically insignificant. To provide an opportunity for valid causal inference, one would have to go behind this observable pattern of seasonal and regional variation, explaining it by means of measurements on those substantive factors that act to make one county or one season different from another in terms of accident risk. A fit approaching 100% is a necessary, but not a sufficient condition for a perfectly specified causal model.

Comparing the models for injury accidents to those for fatal accidents, the explanatory power, as measured by traditional goodness-of-fit statistics, is much higher for injury accidents. This is, however, due exclusively to the fact that the natural amount of random variation is (relatively speaking) much larger for rarer events. Indeed, when due account is taken of the randomness, the models for fatal accidents are seen to explain a larger share of the variation than the models covering all injury accidents. One might only speculate why this is so. It cannot be ruled out that shortfalls in the accident reporting routines, not accounted for by our dummy variables, play an important role in this respect and that, if reporting had been complete also for injury accidents, one would be able to explain as large a share of their systematic variation as is the case for fatal accidents.

5.3. *Models for road users killed*

As is evident from Tables 3 and 4 above, the models explaining, respectively, fatal accidents and road users killed, yield—not surprisingly—very similar parameter estimates. There is one exception: The overdispersion parameter is very much larger in the models explaining death victims. It is not hard to see why. While accidents, for all practical purposes, can be treated as probabilistically independent events, victims are not. An accident may very well involve more than one victim, meaning that the Poisson specification can be regarded only as an ad hoc approximation to the true, victim-generating process. A positive overdispersion parameter must always be expected, no matter how complete a set of independent variables has been included.

The additional overdispersion present in the victims model translates into a reduced goodness-of-fit, as expressed by any one of the five measures defined. In Fig. 3, the randomness component appears to be considerably smaller than in the case of fatal accidents (Fig. 2), also yielding a poorer overall "explanatory power" for all factors taken together. This is, however, clearly a statistical artifact. The scope for random variation is, obviously, at least as large with respect to death victims as it is in the case of fatal accidents. In reality, therefore, we are just about as close to explaining all the explicable (systematic) variation in the death counts as we are in the case of accident counts, the difference being that there is no good yardstick against which to measure explained variation in probabilistically dependent events, such as road deaths.

6. SUMMARY AND CONCLUSIONS

The formulation of (generalized) Poisson regression models for accident counts allows for an interesting opportunity, seldom met with in econometric modelling, of decomposing the total variation in the dependent variable into one part due to normal random (inexplicable) variation, and another part due to systematic, causal factors. It is, in other words, possible to define a yardstick, telling the analyst just how much variation he or she should ideally be able to, or attempt to, explain.

This approach may seem fruitful whenever one wants to explain causally determined, probabilistically independent chance events. It is less well suited for situations in which the events are not truly (objectively) random, or when they are probabilistically dependent. In safety analysis, for example, it seems preferable to work with accidents rather than victims as the dependent variable. Analyses performed on data for Denmark, Finland, Norway, and Sweden suggest that even quite simple Poisson regression models can come very close to explaining almost all the systematic variation in a cross-section/time-series accident data set. When the events analyzed are not independent, however, it is strongly advisable to use a negative binomial rather than a pure Poisson specification, as a certain amount of overdispersion must always be expected in such cases.

The scope for normal random variation is strongly dependent on the size of the unit of observation, as measured by the expected number of events.

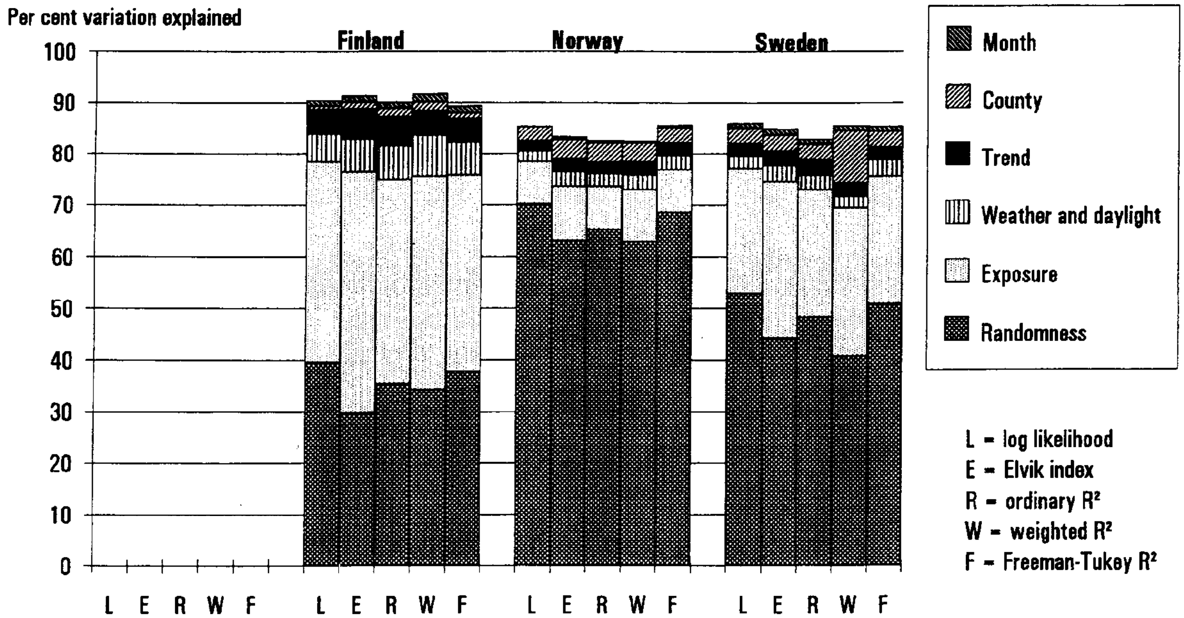

Fig. 3. Variation in road users killed, decomposed by source, according to five different goodness-of-fit measures.

For data sets in which the expected number of events is small—say, always less than 10—a major part of the variation will typically be due to sheer chance. It is useful for the analyst to be aware of the fact that, in such cases, no model should attempt or be able to explain more than the smaller part of the observed variation. When the effects of policy interventions are to be evaluated, it is essential to be able to control for the sometimes very important random component in casualty counts.

The fact that a major part of the observable variation in an empirical data set is attributable to sheer randomness, does not, however, imply that causal factors are in any sense unimportant. The systematic, causal factors determine the expected number of casualties around which the observed variation in casualty counts is centered. By bringing sufficiently large changes to bear on these systematic factors, one would, in principle, be able to alter the long-term casualty count by any positive ratio conceivable.

Inference concerning the partial causal effects of the respective systematic factors at work is subject to a number of pitfalls. Coefficient estimates are derived on the basis of the multiple correlation structure present in the data set and on the model specification chosen. The causal interpretation of these coefficients rests crucially on whether the chosen statistical model corresponds to some true, casualty-generating causal process—an assumption that is ultimately untestable and axiomatic in nature.

While in reality the expected number of road casualties is obviously dependent on a very large number of variables, only a small set of factors has been considered in our analyses. Causal inferences concerning these few factors are biased to the extent that the omitted explanatory variables are correlated with the independent variables included in the model.

The various road safety measures and risk factors that have been in effect over our observation period—omitted to the extent that they are not captured by the uniformly linear trend term —are, indeed, likely to be correlated with the exposure, weather, and daylight variables. These factors are therefore liable to cause certain biases in the exposure, weather, and daylight coefficients. It might be argued, however, that the weather and daylight variables are necessarily exogenous to the economic, social, and accident-generating process, so that if any correlation exists between these variables and any set of omitted explanatory factors, it must be because the omitted variables are influenced by weather and daylight—or at least by their normal seasonal and geographic pattern of variation (the climate)—rather than vice versa. This amounts to saying that, at worst, the weather and daylight coefficients incorporate, not only the immediate effect on casualty risk of, e.g. rainfall and snowfall, but also, to some extent, the effects of certain economic and social variables having, on account of climatic factors, a clear seasonal or geographic pattern of variation.

Perhaps the most prominent source of error in our analyses is accident underreporting, which varies systematically with the type of accident, the calendar, the geographic location, and hence probably with almost all the independent variables used in our models. For fatal accidents and death victims, however, this source of error is fortunately negligible.

The impact of factors that do not vary over the data set is not detectable through an analysis like ours. Thus, risk or safety factors that apply uniformly to all time periods and all regions within a given country offer no power of explanation, nor do measures that had yet to be introduced at the end of the observation period. In general, since our analyses include practically no accident countermeasures, the possible casualty-reducing effects of such measures are obviously not assessable on the basis of our work. It is well known from other sources that certain measures (e.g. seat belts) may have a significant impact on the accident toll, or at least on that subset of victims or accidents to which they are targeted (e.g. car occupants). In our analyses, however, such measures are all subsumed under the trend term, contributing to its coefficient only to the extent that their effects are noticeable in terms of aggregate accident counts. Risk or safety factors having an effect on only a small subset of accidents are not likely to be traceable in macro data, no matter how strong that effect may be.

With these qualifications, our empirical analyses suggest that, among the various factors behind the systematic variation in road casualty counts, exposure is by far the most important, explaining at least 50% of the systematic variation in fatal accidents and more than 70% in the case of injury accidents. The relationship between exposure and injury accidents appears to be almost proportional, while it is less than proportional in the case of fatal accidents or death victims.

Weather conditions appear to play a fairly important role in the accident-generating process, although in the Nordic countries the direction of impact is somewhat counterintuitive. Other things being equal, fewer injury road accidents and deaths seem to occur under typical winter conditions than otherwise. It cannot be ruled out, however, that this finding is biased by the fact that exposure levels are not perfectly controlled for in our models.

A surprisingly large share of the variation in road casualty counts is explicable in terms of factors not ordinarily within the realm of traffic safety policy. In view of this observation, and of the fact that injury accidents appear to vary almost proportionately with the gasoline sales, it may seem unlikely that very substantial reductions in the aggregate accident toll can be achieved without a decrease in the one most important systematic determinant: the traffic volume.

Acknowledgements—This research was commissioned by the Nordic Council of Ministers and is fully documented in report Nord 1993:35, available from the Nordic Council of Ministers, Store Strandstræde 18, DK-1255 Copenhagen, Denmark. Additional support was received through the PROSAF program of the Royal Norwegian Council for Scientific and Industrial Research (NTNF). The authors are also indebted to Erik Biørn, Rune Elvik, Marc Gaudry, and Tore Vaaje, as well as to three anonymous referees, for their constructive suggestions, criticism and ideas, and to Jan Erik Lindjord for his computing assistance.

REFERENCES

Adams, J. G. U. Smeed's law: Some further thoughts. Traffic Engineering and Control 28(2):70–73; 1987.

Bickel, P. J.; Doksum, K. A. Mathematical statistics: Basic ideas and selected topics. San Francisco: Holden-Day; 1977.

Borger, A. Underrapportering av trafikkulykker. Oslo: Institute of Transport Economics; note 975, 1991.

Brännäs, K.; Johansson, P. Time series count data regression. Umeå Economic Studies no 289. Umeå: University of Umeå. 1992.

Brorsson, B.; Ifver, J.; Rydgren, H. Injuries from single vehicle crashes and snow depth. Accid. Anal. Prev. 20:367–377; 1988.

Cameron, A. C.; Trivedi, R. K. Econometric models based on count data: comparison and applications of some estimators and tests. Journal of Applied Econometrics 1:29–53; 1986.

Cartwright, N. Nature's capacities and their measurement. Oxford: Clarendon Press; 1989.

Finnish Roads and Waterway Administration: Liikenneonnettomuustilastojen edustavuustutkimus 1982. Osa IV: Pääraportti, yhteenveto erillistutkimuksista. (English summary: "The sample match of the traffic accident statistics in 1980") Helsinki: Tie ja vesirakenushallitus liikenetoimisto, Liikennevakuutusyhdistys, Kehittämistoimisto Oy Erg AB; 1982.

Freeman, M. F.; Tukey, J. W. Transformations related to the angular and the square root. Annals of Mathematical Statistics 21:607–611; 1950.

Fridstrøm, L. In favor of aggregate econometric accident models. Quebec: 6th International Conference on Travel Behaviour; 1991.

Fridstrøm, L. Causality—is it all in your mind? An inquiry into the definition and measurement of causal relations. In: Ljones, O.; Moen, B.; Østby, L., editors. Mennesker og modeller—livsløp og kryssløp. Oslo: Central Bureau of Statistics; Social and Economic Studies 78; 1992.

Fridstrøm, L.; Ingebrigtsen, S. An aggregate accident model based on pooled, regional time-series data. Accid. Anal. Prev. 23:363–378; 1991.

Gothenburg City Planning Office. Hur många skadas i trafiken egentligen? En analys av skillnaden mellan polisens och sjukhusens trafikskaderapportering i Göteborg 1983. Gothenburg: Trafikdata 2/86; 1986.

Gourieroux, C.; Monfort, A.; Trognon, A. Pseudo maximum likelihood methods: Theory. Econometrica 52:681–700; 1984a.

Gourieroux, C.; Monfort, A.; Trognon, A. Pseudo maximum likelihood methods: application to Poisson models. Econometrica 52:701–720; 1984b.

Greene, W. H. LIMDEP. New York: Economeric Software, Inc.; 1990.

Kulmala, R. Accident models for rural highway junctions. Proceedings of Seminar K, 19th PTRC Summer Annual

Meeting. London: The Planning and Transport Research and Computation International Association; 1991:171–180.
Larsen, C. F. (editor). Personskader opstået ved trafikulykker, behandlet på skadestuen, Odense Sygehus. Odense Hospital: UlykkesAnalyseGruppen; 1989.
Maycock, G.; Hall, R. D. Accidents at 4-arm roundabouts. Crowthorne, Berkshire: Transport and Road Research Laboratory; TRRL Laboratory Report 1120, 1984.
McCullagh, P.; Nelder, J. A. Generalized linear models. London: Chapman and Hall; 1983.
Nedland, K. T.; Lie, T. Offisiell statistikk over trafikkulykker er ufullstending og skjev. Oslo: Institute of Transport Economics; note 786, 1986.
Papineau, D. For science in the social sciences. London: Macmillan Press Ltd; 1978.
Papineau, D. Probabilities and causes. Journal of Philosophy LXXXII (2):57–74; 1985.
Rideng, A. Transportytelser i Norge 1946–1992. Report 187. Oslo: Institute of Transport Economics. 1993.
Salmon, W. C. Scientific explanation and the causal structure of the world. Princeton: Princeton University Press; 1984.
Suppes, P. A probabilistic theory of causality. Amsterdam: North-Holland; 1970.
Thulin, H. Trafikolyckor och trafikskadade enligt polis, sjukvård och försäkringsbolag. Linköping: Swedish Road and Traffic Research Institute; VTI-meddelande 547, 1987.

APPENDIX

A. GOODNESS-OF-FIT MEASURES FOR GENERALIZED POISSON REGRESSION MODELS

Traditional measures of goodness-of-fit describe how much "variation" in the dependent variable is "explained" by the model estimated, as compared to the total variation present in the data set. *Variation* is usually taken to mean "sample variance", or "the sum of squared deviations from the sample mean".

In most situations, however, goodness-of-fit measures are of limited value to the researcher or policy maker, since it is not at all clear that the "best" model is the one that provides the best fit, nor is it clear just how good a fit should be, in order for the model to be judged acceptable. The maximally obtainable fit would depend on the level of measurement and distribution of the dependent variable, in particular on the true structure of the error-generating process, which introduces random variation into the data set. There is no point in trying to explain the purely random noise. On the contrary, the aim of the analysis is to explain all systematic variation, i.e. all variation except the part due to sheer randomness. But since we normally do not know how much random disturbance there should be, there is no yardstick against which we can measure a given goodness-of-fit in order to judge whether or not a given model explains all the systematic variation there is to explain.

In this context, the Poisson regression model, when appropriate, offers some rather interesting opportunities. In a perfectly specified and estimated Poisson regression model, i.e. one in which we have actually found the true value of the Poisson parameter determining the distribution of the dependent variable (e.g. the accident intensity) at each sample point, we also know how much deviation from those true, expected values is to be expected due to purely random variation. Having "explained" all the systematic variation, we have, in other words, a way to predict even the random one. The variance of a Poisson variable is equal to its mean. Unlike the situation in most other regression models, it is, in principle, possible to compute an optimal fit, determined by the amount of (purely random) variation that would be present in the perfectly specified model. Any model providing a higher explanatory power than this would have to be discarded as underdispersed (overfitted), meaning that part of the purely random variation has in fact, through our estimation procedure, been treated and interpreted as systematic.

There are several ways in which one can measure goodness-of-fit. We shall start by examining the familiar coefficient of determination, sometimes referred to as the (squared) multiple correlation coefficient (R^2).

A.1. R^2—the squared multiple correlation coefficient

If we denote by y_i the observations on our dependent variable, by $\bar{y}$ their sample average, by $\hat{y}_i$ the fitted values from some model estimation, and by $\hat{u}_i = y_i - \hat{y}_i$ the residuals, the usual R^2 measure can be written as

$$R^2 = 1 - \frac{\Sigma_i \hat{u}_i^2}{\Sigma_i (y_i - \bar{y})^2} = \frac{\Sigma_i (y_i - \bar{y})^2 - \Sigma_i \hat{u}_i^2}{\Sigma_i (y_i - \bar{y})^2}, \tag{1}$$

where the second term after the first equality sign is interpretable as the residual (unexplained) part of the total sample variation. The remaining part—R^2—is "explained" by the model.

Now, if the y_i are Poisson-distributed with parameters, λ_i (say), the expected value of $\hat{u}_i^2$ is also (approximately) λ_i, meaning that in a perfectly specified and estimated model, the total residual variation would have an expected value approximately* given by

$$E(\Sigma_i \hat{u}_i^2) \approx \Sigma_i \lambda_i = \Lambda \text{ (say)}. \tag{2}$$

A consistent estimate of Λ is

$$\hat{\Lambda} = \Sigma_i \hat{y}_i. \tag{3}$$

Thus, even in a perfectly specified and estimated model the amount of "explained" variation would not exceed

$$P^2 = 1 - \frac{\Sigma_i \hat{y}_i}{\Sigma_i (y_i - \bar{y})^2}. \tag{4}$$

Here, the second term on the right-hand side is interpretable as the random noise part of the total variation, while the remaining part—P^2—is the estimated amount of systematic variation.

Now, since P^2 is in a sense the upper bound on the amount of variation that we would want to or be able to

*In smaller samples one should correct for the degrees of freedom, multiplying by $(n - k)/n$, where k is the number of parameters in the model.

explain, a natural goodness-of-fit measure for the Poisson model could be

$$R_P^2 = \frac{R^2}{P^2} = \frac{\Sigma_i (y_i - \bar{y})^2 - \Sigma_i \hat{u}_i^2}{\Sigma_i (y_i - \bar{y})^2 - \Sigma_i \hat{y}_i}. \tag{5}$$

Note that this measure differs from the ordinary R^2 statistic only in that the amount of normal random variation has been subtracted from the total sample variation appearing in the denominator. The "yardstick" has, in a sense, been adjusted, so as to leave out the inevitable random disturbance from the total sample variation. Only the remaining, systematic variation is of interest for purposes of interpretation.

This adjustment has the interesting consequence that the R_P^2 measure could conceivably exceed 1. Such an instance would, however, have to be interpreted as an indication of overfitting: more variation has been explained than there really is to explain. Certain parts of the explained variation would simply be a misinterpretation of purely random noise, due to spurious correlation rather than to real causal relationships.

Fortunately, the denominator of R_P^2 is fairly robust with respect to the different model specifications, at least as long as the models contain a constant term, since in this case the sum of the fitted values $\hat{y}_i$ will be only marginally different from the sum of observed values y_i. Exact equality will, however, apply only in the case of ordinary least squares (OLS) linear regression.

The usual (and preferred) way to estimate a Poisson regression model is not by OLS, but by maximum likelihood (ML) or some variant thereof. It might be argued, therefore, that R^2 is not the relevant goodness-of-fit measure for Poisson models, and that a measure based on the (log-)likelihood would be more appropriate. By definition, R^2 is maximized by OLS, but not by ML.

To the extent that one wants to use the goodness-of-fit measure as a basis for specification tests, this argument undoubtedly carries a lot of merit. Our concern, however, is to obtain an intuively understandable way of decomposing the total variation present in a sample, into one part explained by our model, and one part not explained. From the unexplained part we would like to be able to subtract the purely random variation that should not under any circumstances be explained, while the explained part should ideally be further decomposable into parts attributable to the different independent variables included in the model.

Moreover, whenever one wants to compare different, non-nested methods of estimation (i.e. methods such that one is not always a special case of another), there is some point in not using a goodness-of-fit measure that is inherently maximized by one of the methods considered. Rather, one needs a measure that can be considered more or less neutral as between the different methods under consideration.

Any such measure is, however, necessarily arbitrary. That is why we present more than one measure. A natural generalization of the ordinary R^2 measure is the weighted R^2.

A.2. The weighted R^2

A Poisson regression model is typically heteroskedastic, in that the variances λ are not equal between different observations i. Indeed, the theoretically optimal linear regression method is one which weights each observation by the reciprocal of its standard deviation, i.e., by $1/\sqrt{\lambda_i}$. Since the λs are unknown, the best one can do in practice (within the class of linear models) is to run a two-stage (or iterative multi-stage) procedure, in which the optimal weights in stage m are approximated by $1/\sqrt{\hat{\lambda}_i^m}$, where $\hat{\lambda}_i^m$ denotes the fitted values calculable from stage $m - 1$. This is tantamount to maximizing the goodness-of-fit measure

$$R_w^2 = 1 - \frac{\Sigma_i \hat{u}_i^2/\hat{y}_i}{\Sigma_i (y_i - \bar{y})^2/\hat{y}_i} = 1 - \frac{X^2}{S_w} \text{(say)}. \tag{6}$$

Note that X^2 is, in fact, the familiar Pearson chi-square statistic, which is minimized through this weighted least squares procedure.

Also, note that under the Poisson assumption, each element $\hat{u}_i^2/\hat{y}_i$ of the sum X^2 has an expected value of approximately one (the weights are approximately variance stabilizing on y_i), so that the Pearson chi-square statistic would be tending approximately towards n, the sample size (more accurately towards $n - k$, the degrees of freedom, where k is the number of parameters estimated). The weighted regression analogue of P^2 can therefore be written

$$P_w^2 = 1 - \frac{n}{\Sigma_i (y_i - \bar{y})^2/\hat{y}_i}, \tag{7}$$

and a weighted goodness-of-fit measure for systematic variation in a Poisson model is given by

$$R_{PW}^2 = \frac{R_w^2}{P_w^2} = \frac{\Sigma_i (y_i - \bar{y})^2/\hat{y}_i - \Sigma_i \hat{u}_i^2/\hat{y}_i}{\Sigma_i (y_i - \bar{y})^2/\hat{y}_i - n}, \tag{8}$$

by analogy to eqn 5.

This measure takes account of the fact that Poisson variates with a larger variance contain less information, and hence should be given a correspondingly smaller weight in the calculation. Equal amounts of information, as measured by the reciprocal of the standard deviation pertaining to each observation, are in a sense given equal weights in the assessment of the model's explanatory power.

This R_{PW}^2 goodness-of-fit measure can, of course, be calculated no matter what method—e.g. maximum likelihood—has been used to estimate the parameters. The maximum likelihood (ML) method, while asymptotically efficient under the Poisson assumption, does not imply maximization of R_{PW}^2 (or R_W^2)—only weighted least squares (WLS) does. The ML method does, however, implicitly take account of the heteroskedasticity, and is therefore unlikely to yield very different results from WLS, except when small values of λ_i predominate in the sample. In such cases the Poisson distribution is heavily skewed, since the range of possible values does not extend below zero. For large λs, on the other hand, the Poisson distribution is approximately normal (Gaussian), in which case WLS is asymptotically efficient and hence almost equivalent to ML (although not algebraically identical).

Note that, because of the weights $1/\hat{y}_i$, the denominators of R_W^2, P_W^2, and R_{PW}^2 are not invariant under different model specifications. They are sensitive to the choice of model for estimating $\hat{y}_i$, especially when the expected values are small, in which case even minor errors of estimation would tend to be inflated through the application of the prescribed weights.

One way to approximate the Poisson to a normal distribution, for small as well as for large values of λ, is by means of the Freeman-Tukey transformation (1950). This transformation is the basis for our third goodness-of-fit measure.

A.3. The Freeman-Tukey R^2

Freeman and Tukey (1950) suggested the following variance stabilizing transformation of a Poisson variable y_i with mean λ_i:

$$f_i = \sqrt{y_i} + \sqrt{y_i + 1}. \tag{9}$$

This statistic is approximately normally distributed with mean

$$\phi_i = \sqrt{4\lambda_i + 1} \tag{10}$$

and unit variance. In other words, the Freeman-Tukey deviates

$$e_i = f_i - \phi_i \tag{11}$$

have an approximate, standard normal distribution. We estimate these deviates by the corresponding residuals

$$\hat{e}_i = \sqrt{y_i} + \sqrt{y_i + 1} - \sqrt{4\hat{y}_i + 1}. \tag{12}$$

An R^2 goodness-of-fit measure for the Freeman-Tukey transformed variables is

$$R^2_{FT} = 1 - \frac{\Sigma_i \hat{e}_i^2}{\Sigma_i (f_i - \bar{f})^2}. \tag{13}$$

Since the Freeman-Tukey deviates have variance one, the maximally obtainable fit in a perfect Poisson model is

$$P^2_{FF} = 1 - \frac{n}{\Sigma_i (f_i - \bar{f}\,)^2}, \tag{14}$$

and the Freeman-Tukey goodness-of-fit measure for systematic variation, analogous to R^2_P and R^2_{PW}, is

$$R^2_{PFT} = \frac{R^2_{FT}}{P^2_{FT}} = \frac{\Sigma_i (f_i - \bar{f})^2 - \Sigma_i \hat{e}_i^2}{\Sigma_i (f_i - \bar{f})^2 - n}. \tag{15}$$

A.4. The Elvik index

Recall the generalization of the Poisson regression model, developed by Gourieroux, Monfort, and Trognon (1984a, b), according to which the Poisson parameter is itself considered random and drawn from a gamma distribution with mean μ_i and shape parameter $\xi = 1/\theta$ (say). In this case the y_i variable can be shown to follow a negative binomial distribution with mean μ_i and variance

$$\sigma_i^2 = \mu_i (1 + \theta \mu_i). \tag{16}$$

We can interpret θ as an overdispersion parameter, indicating a larger empirical variance than the normal random disturbance in a perfectly specified Poisson model, where we would have $\theta = 0$. The higher the overdispersion parameter is, the less systematic variance is actually explained by the model.

Rune Elvik (personal communication) has suggested to us a way to exploit this fact in order to define a relative measure of explained variance. Given that a given compound gamma-Poisson model (named m, say) has been estimated (by ML or a variant thereof), one computes the individual variance estimates

$$(\hat{\sigma}_i^m)^2 = \hat{y}_i^m (1 + \hat{\theta}^m \hat{y}_i^m) \tag{17}$$

and their sum

$$E^m = \Sigma_i \hat{y}_i^m + \hat{\theta}^m \Sigma_i (\hat{y}_i^m)^2 = \hat{\Lambda}^m + \hat{\theta}^m \Sigma_i (\hat{y}_i^m)^2. \tag{18}$$

Let model 0 be the very simplest model estimable, i.e. one with only a constant term and no explanatory, independent variables. In this model the overdispersion parameter ($\hat{\theta}^0$, say) will be larger than in any other model—so large, in fact, that it reflects all the systematic variation there is in the sample. The fitted values $\hat{y}_i$ will, however, be all equal in this model, and obviously very biased estimates of the individual expected values μ_i. The sample variance of the $\hat{y}_i$ will be zero.

To define a yardstick for overdispersion in a gamma-Poisson model, one must therefore choose one fairly trustworthy reference model (i.e. one that makes theoretical sense and exhibits a very small amount of overdispersion), and compute

$$E^{0*} = \Sigma_i \hat{y}_i^* + \hat{\theta}^0 \Sigma_i (\hat{y}_i^*)^2 = \hat{\Lambda}^* + \hat{\theta}^0 \Sigma_i (\hat{y}_i^*)^2, \tag{19}$$

where the $\hat{y}_i^*$ denote the fitted values of the reference model, i.e. our best estimates of the true expected values μ_i (ideally, the true values themselves should have been used).

To judge model m against this yardstick, compute

$$E^{m^*} = \Sigma_i \hat{y}_i^* + \hat{\theta}^m \Sigma_i (\hat{y}_i^*)^2, \tag{20}$$

the total amount of dispersion in model m, and the relative measure

$$R^2_E = 1 - \frac{E^{m^*}}{E^{0^*}} = \frac{(\hat{\theta}^0 - \hat{\theta}^m)\, \Sigma_i (\hat{y}_i^*)^2}{\hat{\Lambda}^* + \hat{\theta}^0 \Sigma_i (\hat{y}_i^*)^2}, \tag{21}$$

where we have defined

$$\hat{\Lambda}^* = \Sigma_i \hat{y}_i^* = E^{m^*}\Big|_{\hat{\theta}^m = 0}. \tag{22}$$

We shall refer to R^2_E as the *Elvik index*. $\hat{\Lambda}^*$ denotes the sum of expected values under the reference model, which we interpret as the minimum amount of dispersion E^{m^*}, consistent with a zero value of $\hat{\theta}^m$. In other words, $\hat{\Lambda}^*$ represents the random noise part of the total variation, while the overdispersion present in a given model m is represented by the second term $\hat{\theta}^m \Sigma_i (\hat{y}_i^*)^2$.

In the special case $\hat{\theta}^m = 0$ (no overdispersion), the Elvik index degenerates into

$$P^2_E = 1 - \frac{\hat{\Lambda}^*}{E^{0^*}}, \tag{23}$$

which is the best fit that one could possibly hope to achieve in a Poisson model.

Now, by analogy to R_P^2, an Elvik index disregarding the normal amount of random noise can be defined by

$$R_{PE}^2 = \frac{R_E^2}{P_E^2} = \frac{E^{0*} - E^{m*}}{E^{0*} - \hat{\Lambda}*} = \frac{(\hat{\theta}^0 - \hat{\theta}^m)\,\Sigma_i(\hat{y}_i^*)^2}{\hat{\theta}^0\,\Sigma_i(\hat{y}_i^*)^2}. \tag{24}$$

A.5. The likelihood ratio goodness-of-fit statistic

Let L^m denote the (maximized) likelihood under model m, and let m and q denote two models (parameter spaces) such that m is a special case of q (i. e. the parameter space under m is contained in that of model q). Assume, as our null hypothesis, that model m is the true model. Then it can be shown (see, e.g. Bickel and Doksum 1977), that, under fairly general conditions (covering the Poisson and negative binomial cases), the likelihood ratio statistic

$$-2\ln(L^m/L^q) = G^2(m,q)\ \text{(say)} \tag{25}$$

is asymptotically chi-square distributed with $k^m - k^q$ degrees of freedom, k^i being the number of parameters under model i. We can use this fact to test model m versus the alternative q.

The most general alternative against which to test is the so-called full (saturated) model—f, say—in which there are as many parameters as there are observations (no degrees of freedom), yielding a perfect fit to the data. To test against this model one computes

$$-2\ln(L^m/L^f) = D^m\ \text{(say)}, \tag{26}$$

which is recognized as the "scaled deviance" measure commonly referred to within the framework of the generalized linear models (GLM) of McCullagh and Nelder (1983). Under model m, D^m has an asymptotic chi-square distribution with $n - k^m$ degrees of freedom.

In the case of Poisson models, the log-likelihood under model m takes the form of

$$\ln L^m = \Sigma_i[y_i \ln \hat{y}_i^m - \hat{y}_i^m - \ln(y_i!)]. \tag{27}$$

Under the full model (f), we have $\hat{y}_i = y_i$, so that

$$D^m = 2\,\Sigma_i[y_i \ln(y_i/\hat{y}_i^m)] - 2(\Sigma_i y_i - \Sigma_i \hat{y}_i^m), \tag{28}$$

where the latter term is quite small provided model m contains a constant term.

For the more general negative binomial specification, we have (see e.g. Cameron and Trivedi 1986)

$$\ln L^m = \Sigma_i\{\ln\Gamma(y_i + 1/\hat{\theta}) - \ln\Gamma(1/\hat{\theta}) - \ln(y_i!) - (1/\hat{\theta})\ln(1+\hat{\theta}\hat{y}_i^m) + y_i \ln[\hat{\theta}\hat{y}_i^m/(1+\hat{\theta}\hat{y}_i^m)]\}. \tag{29}$$

As θ approaches zero, the underlying gamma distribution becomes degenerate, and the negative binomial likelihood converges to the Poisson.

Now, let D^0 denote the scaled deviance of the zero model, i.e. the one with only a constant term and an overdispersion parameter. This deviance incorporates literally all the variation present in the sample. An intuitively obvious way to measure explanatory power would be to relate the scaled deviance of any given model m to that of the zero model:

$$R_D^2 = 1 - \frac{D^m/(n-k^m-1)}{D^0/(n-2)} = \frac{D^0/(n-2) - D^m/(n-k^m-1)}{D^0/(n-2)}, \tag{30}$$

where the deviances are always measured in terms of their degrees of freedom. (For comparability with the Poisson model, we denote by k^m the number of parameters *in addition to* θ.)

Now, under the hypothesis that model m explains all systematic variation, θ equals zero, the Poisson likelihood applies, and the expected value of the deviance is given by

$$E(D^m) = 2\cdot E\{\Sigma_i[y_i \ln(y_i/\lambda_i) - (y_i - \lambda_i)]\} = 2\cdot\Sigma_i\{E[y_i \ln(y_i)] - \lambda_i \ln(\lambda_i)\}. \tag{31}$$

There is, to our knowledge, no closed form solution to this expression. We have, however, proceeded to compute $E[y\ln(y)]$ for a range of values of λ ($e^{-2} < \lambda < e^6$), and fitted a 12th degree polynomial in λ, given by

$$E[y_i \ln(y_i)] - \lambda_i \ln(\lambda_i) = \sum_{k=-6}^{6} \beta_k \lambda_i^k + \varepsilon_i \tag{32}$$

($\lambda_i = e^{-2+i/20}$, $i = 0,1,2,\ldots,160$). As λ_i grows, the difference on the left-hand side approaches one half. In a data set with only large values of λ, the expected value $E(D^m)$ converges, therefore, to n—the sample size. For small values of λ ($\lambda < .5$, approximately), however, the difference is less than one half, while it is larger than one half for moderate values of λ above .5 (see graph by Maycock and Hall (1984, p 61)).*

Now, to purge the deviance of its expected random noise, we estimate $E(D^m)$ by substituting the fitted values $\hat{y}_i^*$ for the true Poisson parameters in eqs (31) and (32), define

$$P_D^2 = 1 - \frac{D_E^m/(n-k^m)}{D^0/(n-2)} \tag{33}$$

and compute (following Kulmala 1991)

$$R_{PD}^2 = \frac{R_D^2}{P_D^2} = \frac{D^0/(n-2) - D^m/(n-k^m-1)}{D^0/(n-2) - D_E^m/(n-k^m)}, \tag{34}$$

where $D_E^m = E(D^m)\Big|_{\lambda_i = \hat{y}^i}$.

A.6. Empirical results

The empirical results obtained by applying these measures to our data set have been presented, in broad terms, in section 5 of the main text (see Figs. 1 through 3).

In the models for injury accidents, the ordinary R^2 measure comes out with values ranging from .62 to .93, depending on the data set (country) and the model specification (i.e. disregarding the models with only a constant term). The weighted R^2 measures range from .72 to .96, the Freeman-Tukey R^2 from .66 to .93, the crude Elvik index from .68 to .94, and the deviance (log-likelihood) R^2 from .66 to .93. By and large, the range of variation is remarkably similar among the five measures.

*Users of the GLIM computer software may calculate an estimate of $E(D^m)$, as given by eqn (31), by using the fitted values $\hat{y}_i$ as estimates of the Poisson parameters λ_i, and summing over all possible values of y_i (see Maycock and Hall 1984). This method is especially useful when the value of the Poisson parameter (the expected number of accidents) is low, i.e. when the possible values of y_i range from zero to 20–30.

The optimal fit, as expressed in terms of the corresponding P^2 measures, are in all cases found to be between .91 and .98, meaning that no matter how well we specify the model, an inexplicable random disturbance component of at least 2% to 9% of the total variation is to be foreseen. This random component is markedly larger in Norway than in the other three countries, simply because the Norwegian units of observation (counties) are smaller (i.e. have fewer casualties on the average), leaving relatively more room for purely random variation.

When we purge the goodness-of-fit measures of this normal amount of purely random noise, measures ranging from .68 to .99 are found. According to the R_P^2 measure, the share of explained systematic variation varies between .68 and .97. The R_{PW}^2 measure varies between .77 and .99, R_{PFT}^2 between .72 and .98, R_{PE}^2 between .73 and .98, and the log-likelihood goodness-of-fit measure R_{PD}^2 between .72 and .98. Again, for one and the same model and data set, the measures—especially the three latter ones—are remarkably similar. In general, the R_P^2 measure comes out somewhat smaller than the rest, while the R_{PW}^2 measure seems to exaggerate the fit slightly, compared to the other four measures. The Freeman-Tukey R^2, the Elvik index, and the scaled deviance (log-likelihood) R^2 appear to be, for all practical purposes, almost equivalent measures of fit.

As we turn our attention to the models for fatal accidents, interesting similarities and dissimilarities (with respect to the injury accident models) can be observed. The crude R^2 measures (i.e. R^2, R_W^2, R_{FT}^2, R_E^2, and R_D^2) are as low as .09 to .12 for Norway, and between .21 and .30 for Denmark and Sweden. This is, however, mainly a reflection of the fact that fatal accidents are less frequent events, leaving relatively more scope for purely random variation. Thus, the optimal (or best obtainable) fit, as witnessed by the P^2 measures, is of the order of 18% to 23% for Norway, and around 40% to 45% for the Danish and Swedish data sets. The share of explained systematic variation (R_{PD}^2, say) is therefore similar to the level found for injury accidents. The models F4, e.g. explain 71% of the systematic variation in Denmark, 81% in Norway, and 88% in Sweden, versus 87%, 84%, and 90%, respectively, for the corresponding models (I4) for injury accidents.

Again, the five sets of goodness-of-fit measures considered are seen to yield fairly similar results, although, as for injury accidents, the measures based on the weighted R^2 are on the high side, while those based on the ordinary R^2 come out rather low. The measures based on the Elvik index are generally very similar to those derived from the log-likelihood. The P_W^2 measure has the undesirable feature of being strongly model dependent, an instability that carries over even to the R_{PW}^2 measure.

Finally, the models for road users killed are seen to explain a strikingly smaller share of the systematic variation than the corresponding models for fatal accidents, despite the fact that the two casualty counts are very strongly correlated, being identical in many cases. This, however, serves to illustrate the importance of independence between events. For events that are not probabilistically independent, the Poisson probability model is obviously less well founded. In this case, a considerable amount of overdispersion must always be expected, translating into a less than perfect fit even after we correct for a normal amount of random Poisson disturbance.

A.7. Synthesis

While, according to traditional measures of goodness-of-fit, our fatal accident models would appear to be very much inferior to the injury accident models, such is not the case when due account is taken of the unequal sizes of the normal random variation in the two cases. This, in our view, illustrates the usefulness of the goodness-of-fit measures developed here. By leaving out the inexplicable from the unexplained variation, one can compute goodness-of-fit measures that are just about equally informative no matter how frequent or rare the events being analyzed. Put otherwise, these measures are able to tell the analyst how far he/she is from explaining all the explicable (systematic) variation there is.

Five measures of explained systematic variation have been considered, of which the one based on the log-likelihood statistic (R_D^2) would appear to be the most "natural" choice, in particular when Poisson or negative binomial maximum likelihood is the method of estimation used. Unlike the other measures, however, the R_D^2 measure is not straightforward to compute, as it depends on the unknown quantity $E[y_i \ln (y_i)]$ (where $y_i \sim \mathcal{P}(\lambda_i)$), whose value must be evaluated for each unit of observation, and summed through the sample. In contrast, the R_P^2, R_{PW}^2, and R_{PFT}^2 measures depend only on the first and second order moments of the observed and fitted values of the dependent variable, or of simple transformations thereof. The Elvik measure R_{PE}^2 is equally simple to compute, although one drawback about this measure is the need to choose and estimate a trustworthy reference model, the fitted values of which must be assumed, with negligible error, to coincide with the true expected values of the dependent variable. This choice of reference model is, of course, ultimately arbitrary.

As a shortcut method to assess the amount of explained systematic variation, we would recommend the Elvik measure R_{PE}^2, except in situations where a trustworthy reference model cannot be found. In such cases, the Freeman-Tukey measure R_{PFT}^2 is a commendable, robust alternative. We advise against using the weighted R^2 measures R_W^2, P_W^2, and R_{PW}^2, as these measures tend to inflate estimation errors affecting the smallest predicted values, and are not invariant with respect to the choice of reference model used to compute normal Poisson disturbances.

Sist utgitte TØI rapporter

Tittel	TØI Rapport
Everyday travel and use of information and telecommunication technology at home: an analysis of Norwegian data.	454/1999
Trafikkpotensialet for norske lufthavner 2001-2010. En vurdering av Flesland, Vigra og Kjevik.	449/1999
Utvärdering av trafiksäkerhets- och miljöreformer	445/1999
Yrkesaktives reiseaktivitet. En analyse av faktorer som påvirker bilordninger, parkeringsmuligheter, omfang og transportmiddelbruk på reiser til/i arbeid i Oslo-området	444/1999
Reisevaner på rutefly 1992-1998	441/1999
Trafikksikkerhet for eldre : Litteraturstudie, risikoberegninger og vurdering av tiltak.	440/1999
Utenlandsk bilturisme i det nordlige Norge 1998	439/1999
Effekter av byspredning på bilhold og bilbruk. En studie av Oslo og Bergen pendlerregioner.	438/1999
Tiltakspakker for kollektivtransport. Evaluering av prosessen rundt tiltakspakkene i Tønsberg og Drammensområdet	437/1999
Daglige reiser på 90-tallet. Analyser av de norske reisevaneundersøkelsene fra 1991/92 og 1997/98	436/1999
Cost-Benefit Analysis of Safety Measures for Vulnerable and Inexperienced Road Users	435/99
Alternative systemer for beregning av engangsavgift på personbiler	434/1999
Samfunnsøkonomiske analyser innen luftfart. Del 1: Veileder for Luftfartsverkets prosjektvurderinger	433/1999
Gå eller sykle? Fakta om omfang, sikkerhet og miljø	432/1999
Flaskehalsproblemer i samferdselsforskningen	431/1999

This report can be ordered from
Institute of Transport Economics
P.O.Box 6110 Etterstad • N-0602 Oslo