Ann Reg Sci (1999) 33:269–287

The Annals of
Regional Science

On the relationship between travel time and travel distance of commuters

Reported versus network travel data in the Netherlands

Piet Rietveld[1], Bert Zwart[1], Bert van Wee[2], Toon van den Hoorn[3]

[1] Faculty of Economics, Vrije Universiteit, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands
[2] National Institute of Public Health and the Environment, P.O. Box 1, 3720BA Bilthoven, The Netherlands
[3] Ministry of Transport, P.O. Box 1031, 3000BA Rotterdam, The Netherlands

Received: November 1996/Accepted: January 1998

Abstract. This paper gives a detailed empirical analysis of the relationships between different indicators of costs of commuting trips by car: difference as the crow flies, shortest travel time according to route planner, corresponding travel distance, and reported travel time. Reported travel times are usually rounded in multiples of five minutes. This calls for special statistical techniques. Ignoring the phenomenon of rounding leads to biased estimation results for shorter distances. Rather surprisingly, the distance as the crow flies and the network distance appear to be slightly better proxies of the reported travel time compared with the shortest network travel time as indicated by the route planner. We conclude that where actual driving times are missing in commuting research the other three indicators mentioned may be used as proxies, but that the following problems may emerge: actual travel times may be considerably higher than network times generated by route planners, and the average speed of trips increases considerably with distance, implying an overestimate of travel time for long distance commuters. The only personal feature that contributes significantly to variations in reported travel times is gender: women appear to drive at lower average speeds according to our data. As indicated in the paper this may be explained by the differences in the car types of male and female drivers (females drive older and smaller cars) as well as higher numbers of stops/trip chaining among women. A concise analysis is carried out for carpoolers. Car-pooling leads to an increase in travel time of some 17% compared with solo drivers covering the same distance. In the case of car poolers, the above mentioned measures appear to be very poor proxies for the actual commuting times.

1. Introduction

In theories of travel behaviour, transport costs play an important role. These costs normally consist of a distance dependent and a time dependent part. For car drivers the distance dependent part relates to the fuel costs and the costs of maintenance, repairs and depreciation (as far as these depend on distance). The time dependent part of car drivers relates to the value of travel time. In many applied cases of transportation research the costs are not known explicitly.

What is usually known instead, are the distance, the travel time or both. When we have a closer look at available data on travel times and travel distances it appears that various problems may occur concerning the measurement of these variables. For example:

- we do not know actual distances travelled, but only distances as the crow flies
- in stead of actual distances travelled we know the 'shortest' network distance between the centroids of the zone of origin and destination.
- when actual travel times are given, drivers may apply rounding routines like roundings in terms of units of 5 min or kilometres.
- reported travel times may include time needed for walking to the car, making the car operational, finding a parking place and walking to the final destination.

Some of the measurement problems have received attention in the literature. For example, Nordbeck (1964) addresses the relationship between shortest network distance and distance as the crow flies. He finds detour factors of around 1.20–1.25. The problem of determining the length of trips that remain within a zone has been addressed by Warnes (1972), and Rich (1980). The literature on these measurement problems is rather thin. There is some relationship with literature on distance cognition (see for example Säisä et al. 1986; Péruch et al. 1989). However, in the distance cognition literature it is not the perception of length of trips actually travelled that is studied, but the perception of the length of trips according to cognitive maps.

In the case of the present paper we investigate these measurement problems based on a data set with rather rich data so that we are able to compare these data. More in particular we want to investigate the following questions:

- to what extent can distances as the crow flies be considered as good proxies for distances travelled via actual road networks?
- to what extent can travel times be explained by travel distances?
- is there any way of correcting reported data for rounding?
- do personal features of drivers have an impact on reported travel times?
- what are the consequences of car-pooling on travel times?

The structure of the paper is as follows: In Sect. 2 we introduce and discuss some relevant concepts of distance and travel time of commuters. Relationships between the concepts are formulated in Sect. 3. Section 4 is devoted to a discussion of estimation issues. After a short discussion of the data set used (Sect. 5) empirical results for solo drivers are given in Sect. 6. This is followed by a concise analysis of travel times of car-poolers (Sect. 7). Section 8 concludes.

2. Definitions of travel distance and travel time variables

In our analysis we will use a list of concepts reported below. It will become clear that this list is partly determined by the special data set available. With other data sets we might end up with (slightly) different definitions, but it will not be difficult to apply the necessary adjustments.

We will consider commuting trips of car drivers in the Netherlands. The following concepts will be used:

dc: distance as the crow flies, between the points of gravity of the zone of origin and the zone of destination of a trip.
tn: shortest travel time between the points of gravity of the zone of origin and the zone of destination in road network, determined by means of a route planner on the basis of minimization of travel time
dn: travel distance of the trip in road network associated with the shortest travel time *tn*
tr: actual travel time as reported by respondent.

The difference between the concepts *dc*, *tn*, and *dn* on the one hand and *tr* on the other hand is that *tr* is based on the actual route chosen by a respondent, while the others are based on a route planner. Another difference is that *tr* is from a specific point in a zone to a specific point in another zone; the other concepts relate to the points of gravity of the zones of origin and destination.

A further relevant difference is that in the route planner a certain speed is assumed on the various segments, whereas the driver may use speeds that differ from this speed, for example because his speed behaviour is different, or because he drives during peak hours so that speeds cannot be freely chosen.

Another point that deserves our attention is that the actual route chosen is not necessarily the one leading to the shortest travel time. For example, when drivers have an opportunity to use an express way allowing high speeds, but implying a longer distance, and a route that is longer in travel time, but shorter in distance travelled, it is not clear a priori which of the two routes will be chosen.

Also we should take into account the possibility that the travel time reported by the driver includes elements not present in the time computed by the route planner: time needed to search for a parking place, time needed for walking from the parking place to the final destination, and different stops on the way from origin to destination.

Another factor could be due to perception. For example, travel time during peak hours might be perceived differently from a time period with no congestion on the network.

A final problem is that when drivers report their travel time they often apply a rounding of figures so that the majority of the times reported are in units of 5 min.

Thus there are several reasons why the four concepts mentioned above may lead to rather different types of outcomes. These concepts are further clarified in Fig. 1.

3. Formulation of model for proxies of actual travel time

Consider the case that we have observations for the commuting trips of a number of individuals:

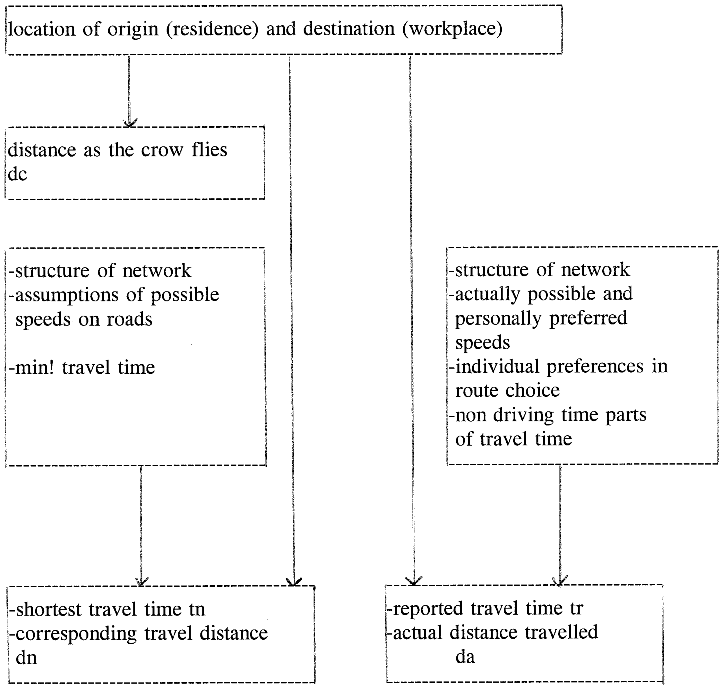

Fig. 1. Schematic presentation of relationships between distance and time concepts.

- *dc*: distance as the crow flies (from centroid to centroid)
- *tn*: shortest travel time in network according to route planner (from centroid to centroid)
- *dn*: travel distance associated with *tn*
- *tr*: reported travel time
- features of the individual
- network structure

Our aim is to formulate equations that allow us to use *dc*, *dn*, and *tn* as proxies for actual reported travel time *tr*. We will use in addition the (unobserved) actual distance *da*.

The chain of relationships to be discussed below is presented in Table 1.

We start with the definition:

$$tr = tr_0 + 1/s \cdot da \tag{1}$$

where tr_0 is the non-driving time in the total travel time.

The speed s relates to the average speed during the commuting trip.

The speed depends on the time of the day (in case congestion is relevant) and on the types of roads available in the network. Since with long trips the share of the trip that takes place via express ways tends to be higher, the speed will depend on the length of the trip.

In addition, as explained in Jorgenson and Polak (1993), speed will depend on the value of time (which in its turn depends on income), on perceived risk

Table 1. Relationships between reported travel times and proxies (x represents other relevant variables).

Reported travel time as a function (direct or indirect) of distance indicators:

$$\left.\begin{array}{l}\left.\begin{array}{l}(3)\ \ tr=f(da,x)\\(4)\ \ da=g(dn,x)\end{array}\right\}\ tr=fg(dn,x)\\(5)\ \ dn=h(dc,x)\end{array}\right\}\ tr=fgh(dc,x)$$

Reported travel time as a function of network travel time:

(6) $tr = k(tn, x)$

of accidents, and on the risk of being fined because of speeding. These authors also find that women tend to drive slower than men. Research by Rouwendal (1996) on energy efficiency of car use (which depends among others on speed) shows that various personal features like employment status and age play a role: a somewhat surprising result is that older drivers are less energy efficient. No significant effect was found for income. Additional (and may be partly contradictory) results are found by Rienstra and Rietveld (1996) who report that income does have a significant positive impact on speeds. The same holds true for the maximum possible speed of the car driven (on highways). Everything else equal, women tend to drive at equal speeds compared with men. However, since women are shown to use slower cars, there is an indirect effect of gender on speed. Older drivers tend to drive at lower speeds. Thus we arrive at the following formulation for speed:

$$s = f(da, \text{personal features, network structure, time of day}) + \varepsilon \tag{2}$$

where ε represents a stochastic term to represent various types of errors.

After substitution of (2) into (1) we arrive at a relationship between the reported travel time tr and the actual travel distance, where also the other factors in (1) and (2) play a role:

$$tr = f(da, \text{personal features, network structure, time of day}, tr_0) \tag{3}$$

This the first equation mentioned in Table 1. However, with the present data set we do not know the actual distance da; instead of this we do know two related distance concepts: dc, the distance as the crow flies; and dn, the distance associated to the shortest route (assuming minimization of distance). Therefore we formulate relationships where dc and dn serve as proxies for da.

For the relationship between da and dn we would have:

$$da = g(\text{personal features, network structure}, dn) + \varepsilon \tag{4}$$

where personal features are added because they may lead people to choose routes that do not correspond to the shortest travel time route. Note that da may be either shorter than dn (if people prefer to take the shortest route in terms of distance, rather than in terms of travel time), but it may also be longer (some people may dislike driving on particular types of roads or may intentionally make a detour to pick up car-poolers or to bring children to school).

Of course, if people would choose the fastest route we would have as a special case:

$$da = dn + \varepsilon, \tag{4'}$$

where ε would account for various errors. Note for example that dn is the distance between the points of gravity of the zones of origin and destination, whereas da is the actual distance between two specific points in these zones.[1]

In addition, the relationship between dn and dc can be specified as:

$$dn = h(\text{network structure, } dc) + \varepsilon \tag{5}$$

A simple representation would be (if differences in network structure would be negligible among the observations):

$$dn = b \cdot dc + \varepsilon \tag{5'}$$

where the detour factor b is of course higher than one, and where this factor might be dependent on the distance dc (smaller detours with longer trips).

Substitution of these functions into each other finally leads to a formulation where, (as also shown in Table 1):

A. tr depends in a non linear way on dn, and on personal features, time of day and network structure
B. tr depends in a non linear way on dc, and on personal features, time of day and network structure

A last possibility, as also shown in Table 1, would be to relate tr directly to the travel time according to the network:

C. tr depends on tn as follows:

$$tr = h(\text{personal features, time of day, network structure, } tn) + \varepsilon \tag{6}$$

When the driver would take the fastest route and drive exactly the speeds used in the network algorithm then this equation can be simplified as:

$$tr = tn + \varepsilon, \tag{6'}$$

where ε is due to the fact that the exact origin and destination of the trips do not coincide with the centroids of the zones of origin and destination as assumed in the network algorithm. However, when the driver would take the fastest route but drive at speeds different from the network algorithm speed the more general formulation above would apply. This formulation also applies when the driver would not take the shortest travel time route.

[1] Note that heteroscedasticity may be a problem here. For short distances coincidental features of a network may lead to large errors. This would imply errors to increase with decreasing distances. On the other hand, one might also have larger errors with larger distances because the error's variance given the sum of two links is the sum of the error's variances for individual links (assuming zero covariance).

Bovy and Stern (1990) discuss a number of issues in the perception of travel time, including the influence of traffic obstacles such as traffic lights and turns. They mention that short times tend to be underestimated, whereas long times are overestimated. However, in the present context, where respondents report about their daily commuting trips this part of the perception issue is most probably not so serious. Systematic misperceptions of their travel time would lead to systematic early or late arrivals at the place of work, which most drivers (or their colleagues) would notice at some stage. There is, however, another problem related to reported travel times which will be discussed in the next section.

4. Estimation issue: Dealing with rounding

The large majority of the reported travel times are rounded as multiples of 5 min. Ignoring this problem by using standard estimation techniques would imply a risk that biased results are found. Therefore we propose to deal with the rounding issue explicitly. This can be done by explicitly introducing the actual travel time *ta* in addition to the reported travel time *tr*. Because of the rounding problem we cannot directly use the equations developed in the preceding section to *tr*; in stead, they will be applied to the unobserved variable *ta*. One must be aware that *ta* is a continuous variable, whereas *tr* is an integer. Thus, even when rounding in multiples of five would not be an issue, rounding as such would still be applied. The difference is of course, that rounding to the nearest integer will have much smaller impacts on the errors so that is can be safely ignored in most cases.

Consider the case of an actual travel time *ta* equal to 17.89 min. We allow the respondent two ways to report this outcome: either in terms of the nearest integer (18 min), or the nearest multiple of 5 (20 min). We also allow that the probability of rounding to a multiple of 5 is higher when the actual travel time is nearer to such a multiple (compare 17.89 with 19.39).

Let $A(i)$ denote the interval around an integer i:

$$\begin{aligned} A(1) &= [0, 1.5] \\ A(i) &= (i - 0.5, i + 0.5] \quad (i > 1) \end{aligned} \tag{7}$$

We assume that rounding never leads to a zero reported travel time (thus an actual travel time of 2.3 min will be rounded as either 2 min or 5 min). Then by using only one parameter s ($s < 1.0$) to represent the rounding procedure we postulate that the probability of rounding to a multiple of five can be formulated as shown in Table 2.

Let $P(ta \in A(i))$ denote the probability that the (unknown) actual travel time is in the interval $A(i)$.

Thus a reported value *tr* of say 10 min is the result of probabilities $P(ta \in A(i))$ for $i = 8, \ldots, 12$ of an actual value *ta* in the intervals $(7.5, 8.5], \ldots (11.5, 12.5]$ with weights $s^2, s, 1, s, s^2$. For a reported value of 5 min summation takes place for all intervals from $[0, 1.5]$ to $(6.5, 7.5]$. When i is not a multiple of 5 the probability of an observation equal to i is $P(ta \in A(i))^*(1 - s)$ when i is a neighbour of a multiple of 5, and $P(ta \in A(i))^*(1 - s^2)$ when i is not a neighbour.

Table 2. Probability of rounding actual travel times *ta*.

Interval of *ta*	Probability of rounding to nearest positive multiple of five	Probability of rounding to nearest positive integer
[0, 1.5]	s^4	$1 - s^4$
(1.5, 2.5]	s^3	$1 - s^3$
(2.5, 3.5]	s^2	$1 - s^2$
(3.5, 4.5]	s	$1 - s$
(4.5, 5.5]	1	1
(5.5–6.5]	s	$1 - s$
(6.5–7.5]	s^2	$1 - s^2$
(7.5–8.5]	s^2	$1 - s^2$
etc.		

Finally we still need an expression for $P(ta \in A(i))$. We will assume here that the unobserved dependent variable *ta* is an additive function of independent variables X plus a normally distributed error term ε with zero mean and standard deviation σ, which is truncated[2]:

$$ta = X\beta + \varepsilon \tag{8}$$

where and σ are parameters to be estimated. The explanatory variables X are those mentioned in equation (6).

Then:

$$\begin{aligned} P(ta \in A(i)) &= P(i - .5 < ta \Leftarrow i + .5) \\ &= \{\Phi[(i + .5 - X\beta)/\sigma] - \Phi[(i - .5 - X\beta)/\sigma]\}/\Phi[X\beta/\sigma] \end{aligned} \tag{9}$$

where Φ is the standard normal distribution function (zero mean, unit variance). The denominator in this function reflects the truncation in the distribution of ε. It is included to take into account the problem that the model may predict negative values for *ta*. Whether this will really occur depends on the value of the σ parameter. When σ is large there is a non-negligible probability that the model predicts a negative value for *ta*. For small values of σ, this probability will be very small. Note that for small values of σ the denominator in (9) approximately equals 1. These equations suffice to formulate the likelihood function and carry out a maximum likelihood algorithm.

5. Description of data

The data have been collected in the context of a study of Van Wee (1995) on the effects of relocation of employers on the locational behaviour of workers: different combinations of changes in jobs and residences are possible (see also Zax 1991).

[2] The error term is truncated in order to ensure that *ta* is positive.

The relocation considered concerns various divisions of the Dutch Ministry of Transport where three decentralized locations (one in The Hague and two in Dordrecht) were replaced by one central location in Rotterdam. The location in Rotterdam is in between the two other locations, the distance being some 20 km to both of them.

The resulting data set is rather rich in the sense that a good number of topics are covered. About 300 workers answered the questionnaire. When we consider commuting distances of car drivers we have 209 useful observations (about half of these relating to the situation before the relocation (1989), and the rest relating to the situation after the move (1993).

In the next section we will give an analysis of the data of 132 trips of solo drivers, followed by a concise analysis of the trips of car-poolers.

6. Empirical results for solo drivers

6.1. General

In Table 3 we give some descriptive statistics for the various concepts. At this stage of analysis data on car-poolers will not be included in the data set.

The mean network distance of commuters *dn* is 25.1 km. The observations range from 0 (when origin and destination of the commuting trip are in the same zone) to 100.0 km. For the *dc* and *tn* distributions similar shapes are found. Rounding clearly influences the shape of the distribution of reported travel times *tr*. The coefficient of variation is clearly lower for the travel times (0.50–0.55) compared with the travel distances (0.70–0.77). The reason is that (as we will see in the next analyses) the average speed of longer trips is higher.

The detour factor, defined as the network distance divided by the distance as the crow flies, following from this table is about 1.40 (25.1/17.9). For the above median trips it is 1.38, whereas for the below median trips it is 1.50[3]. The average speed for this sample of trips is about 62 kmph according to the network algorithm. Note, however, that the reported travel times are considerably higher. There is a clear gap between the means of *tn* and *tr*. According to reported travel times the average speed of commuters is only 50 kmph. In another recent study for the Netherlands (BGC 1996) a similar underestimate was found for travel times. There are several explanations for this gap.

First, the reported travel times possibly include some non-driving time components of travel times. Second, most commuters make their trips during rush hour so that speeds may be lower than assumed in the route planner.

[3] This computation of the detour factor is the weighted average of detour factors of individual trips, the weights being the *dc* distances. In case one would take the unweighted distances we would arrive at an average detour factor of 1.50, with values of 1.56 for below median trips and 1.43 for above median trips.

Note that in a network consisting of a fine meshed rectangular lattice the maximum detour factor for any trip would be 1.41 (the square root of 2). As indicated by Norbeck (1964) and Warnes (1972) the mean detour factor would of course be closer to 1.00 (about 1.2). There are three main reasons that in the present case we find a higher average detour factor. First, the network is not so fine meshed. Second, particular links are missing (for example due to physical boundaries such as river crossings). Third, the network consists of roads with different speeds. Since the *dn* distances are based on the shortest routes in terms of travel time it is plausible that detours are made to save travel time.

Table 3. Descriptive statistics of travel times and distances of car using commuters in the Netherlands (solo drivers).

	dc distance as crow flies (km)	*dn* distance according to route planner (km)	*tn* travel time according to route planner (min)	*tr* reported travel time (min)
Mean	17.9	25.1	24.4	30.3
Standard deviation	13.7	17.5	13.4	15.1
Coefficient of variation	0.77	0.70	0.55	0.50
Minimum	0	0	0	5
25% observation	8.02	13.5	16.3	20
Median	14.1	22.7	21.2	30
75% observation	23.8	33.9	30.5	40
Maximum	75.1	100.0	70.3	90

Table 4. Correlations between travel distances and travel times

	dc	*dn*	*tn*	*tr*
dc	1.000	0.966	0.947	0.798
dn	0.966	1.000	0.930	0.803
tn	0.947	0.930	1.000	0.751
tr	0.798	0.803	0.751	1.000

Third, commuters may not choose the shortest route as assumed in the route planner.

The mutual correlations between these variables are presented in Table 4.

The correlations between *dc*, *dn* and *tn* are high: they range from 0.93 to 0.97. It is interesting to note that the correlations for the reported travel times are clearly lower (they range from 0.75 to 0.80). This lower correlation may be due to the rounding problem mentioned above, but also other factors may play a role. At this stage of analysis there is no indication that the network travel time (*tn*) is a better proxy for the actual reported travel time than the distance based indicators (*dc* and *dn*); rather, the reverse seems to be true.

In this study we will especially focus on reported travel times as an endogenous variable. Before this analysis we will first shortly consider the relationships between the other distance and travel time variables. We specify the relationships in such a way that they are piece-wise linear (see Fig. 2.)

In order to deal with the piece-wise linearity we introduce variables Z_1 and Z_2 and a cutting point B on the scale for X as follows:

$$\begin{aligned} Z_1 = {} & X && \text{if } X < B \\ & B && \text{if } X > B \\ Z_2 = {} & 0 && \text{if } X < B \\ & X - B && \text{if } X > B \end{aligned}$$

The value of B can be chosen on various grounds. In our case we take the median value of X as the value for B.

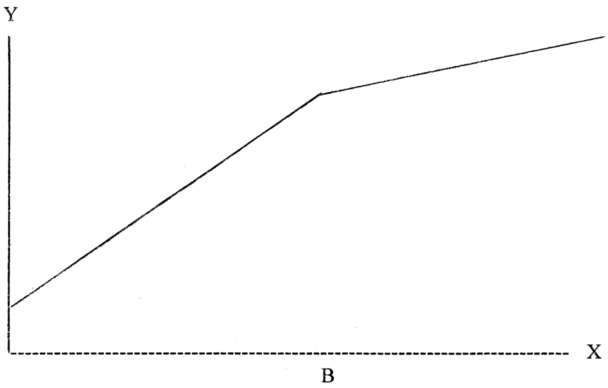

Fig. 2. Piece-wise linear relationship between X and Y.

The following result is obtained for the network distance *dn* when it is related to the distance as the crow flies *dc* (standard errors in brackets, for reasons mentioned in Sect. 3 we take into account the possibility of heteroscedasticity):

$$\underset{(0.18)}{dn = 0.241} + \underset{(0.022)}{1.335^*} \text{ first part } dc + \underset{(0.036)}{1.266^*} \text{ second part } dc + \varepsilon, \tag{10}$$

where ε is a normally distributed error with mean equal to zero and variance equal to $0.241^*(dn+1)^{1.55}$. The value 1.55 appears significantly higher than zero, which implies that heteroscedasticity indeed is present.[4] According to this estimation, the constant term is very small (not significant). For shorter distances the detour is slightly larger, but the difference between the short distance factor (1.34) and the long distance factor (1.27) is not significant in the present case.

In a second step we now explain the network time *tn* by the network distance *dn*:

$$\underset{(0.71)}{tn = 4.21} + \underset{(0.054)}{1.078^*} \text{ first part } dn + \underset{(0.019)}{0.600} \text{ second part } dn + \varepsilon \tag{11}$$

where ε is a normally distributed error with mean equal to zero and variance equal to $0.241^*(tn+1)^{0.293}$. The value 0.293 is significantly higher than zero, implying again heteroscedasticity. Equation (11) indicates a clear non-linear relationship between network time and network distance. The constant term is also higher. This seems to suggest at least three speed regimes: very low speeds at short distances when the driver is near to his origin and/or destination

[4] The conclusion is that of the two countervailing factors mentioned in Sect. 3 the second one appears to be dominant: the larger the distance, the larger the error's variance. We have used $(dn+1)$ rather than dn in equation (10) to avoid problems with zero values for *dn*.

leading to a constant term of about 4 min per trip, a speed of some 56 kmph (60/1.08) for parts of the trip somewhat further away, and 100 kmph (60/.60) for the parts of the trips longer than the median value of 22.7 km. These values obviously depend on the average shares of the various types of roads in the total trip, each with its own speed regime.[5]

6.2. Estimations based on reported travel times

Rounding of travel times appears to be pervasive. Of the 132 travel times reported by solo car drivers 130 (98.5%) are a multiple of 5; this is obviously much higher than the share of 20% one may expect when rounding does not take place (the two non rounded observations of solo drivers are: 16 min and 18 min). Thus there is much reason to use the maximum likelihood estimation technique presented in Sect. 4.

Following the structure of Table 1 in a first specification for an analysis of reported travel time *tr* we make it a function of each of the three other commuting cost indicators: *dc*, *dn* and *tn* (see Table 5).

For the distance based specifications (*dc* and *dn*) we find that speeds are higher at longer distances. For example, the coefficient of 0.607 for *dn* indicates that drivers attain speeds of $60/0.607 = 99$ kmph at the parts of trips longer than 25 km. This is a reasonable outcome since the maximum speed on highways is either 120 or 100 kmph in the Netherlands. The average speed at the shorter part of the trips is 60 kmph.

The results of this table also confirm the statement in Sect. 2 that for *dc* and *dn* we expect a non-linear relationship with reported distances.

For *tn* there are no reasons to expect a non-linear relationship according to Sect. 2. This is confirmed in a statistical sense in Table 5 since the gain in log-likelihood compared with a kinkless specification is not significant. As we observed in Table 1, the reported travel times are almost 25% higher than the trave times according to the route planner. Although the constant term of 2.6 min in Table 5 is not significant in the estimation of *tr*, it explains part of the

[5] Another approach would be to apply a Box-Cox transformation. This transformation allows a flexible non-linear shape. The general formulation is:

$$tn = b0 + b1[(dn^a - 1)/a]$$

When the parameter a equals 1 the relationship between *dn* and *tn* is linear, when a equals 0 the relationship is logarithmic.

The estimation results for the parameters are:

parameter	estimate	standard error
$b0$	2.98	1.10
$b1$	1.26	0.206
a	0.834	0.053

The value for a is significantly different from both 1 and 0. This means that the curve is neither linear nor logarithmic, a conclusion consistent with the piece-wise linear specification. Applications of the Box-Cox specification to the other equations estimated in this paper lead to results that are consistent with the piece-wise linear results in a comparable way.

Table 5. Estimation results for reported travel times as dependent variable (rounding taken into account).

	Independent variable is *dc*	Independent variable is *dn*	Independent variable is *tn*
Rounding parameter *s*	0.987 (0.0088)	0.987 (0.0088)	0.987 (0.0088)
Constant	8.92 (2.41)	7.96 (2.39)	2.62 (4.14)
Slope for below median observations of independent variable	1.435 (0.23)	0.993 (0.14)	1.242 (0.24)
Slope for above median observations of independent variable	0.762 (0.082)	0.607 (0.078)	0.769 (0.13)
log-likelihood	−274.32	−273.34	−287.33
Gain in log-likelihood compared with specification without kink	3.15 (significant)	2.37 (significant)	1.38 (not significant)

gap. In combination with the high slope for shorter trips it appears that the (relative) gap between *tr* and *tn* is largest for shorter distances. Thus a tendency can be observed that the route planner does a better job as a predictor of actual travel time for longer distances compared with shorter distances.

The estimation results support the need of adding a constant term in the travel time estimations in the case of *dn* and *dc*. This may be an indication that reported travel times include time needed to walk to and from parking places. However, these constant terms most probably are also influenced by other factors such a misspecification for trips at very short distances (it may well be that at very short distances lower speeds would apply than implied by the slope at the left side of the link).

An important result of this table is also that the crude measure *dc* performs equally well as a proxy for reported distances compared with the more refined concept of *dn*. In addition it is striking that the travel times based on the shortest route algorithm perform slightly less favourable as a proxy for actual travel times compared with the other variables (note also that in Table 3 the correlation coefficient between *tn* and *tr* is the lowest value found).

The estimation result strongly supports the importance of taking account of the rounding problem. The value of the rounding parameter s implies a probability of 0.987 that respondents with an actual travel time between 23.5 and 24.5 min would round this value as 25 min.

If we would ignore the rounding (and truncation) issue the following results would be found (see Table 6).

When we compare these results with those in Table 5 we find that ignoring the rounding (and truncation) problem leads to an upward shift in the constant term (the range shifts from 3–9 to 7–11 min) and a decrease in the slope for shorter distances of 10 to 20%. For the longer distances the slopes are almost identical. Thus ignoring the rounding problem leads to biased estimates for the lower range of the dependent variable. For the higher range it can safely be ignored.

Finally, we will further investigate the relationship between reported travel times *tr* and network travel times *tn* by including additional explanatory variables. Since we found in Table 5 that a linear specification is adequate we

Table 6. Estimation results for reported travel times as dependent variable (OLS)

	Independent variable is *dc*	Independent variable is *dn*	Independent variable is *tn*
Constant	10.79 (2.23)	9.96 (2.19)	6.83 (3.51)
Slope for below median observations of independent variable	1.29 (0.203)	0.897 (0.125)	1.03 (.203)
Slope for above median observations of independent variable	0.768 (0.076)	0.613 (0.064)	0.782 (0.091)
R^2	0.649	0.653	0.567

take this as a starting point. The following additional explanatory factors have been used:[6]

– year of observation: (observations relate to the years 1989 and 1993). This variable is added to take account of differences in the congestion level between these years, and differences in the location of workplaces.
– age of respondent (dummy = 1 for drivers older than 45 years)
– income level of respondent (dummy = 1 for income higher than dfl 51,000 per year)
– gender of respondent (dummy = 1 for female drivers).
– location of the worker to take into account network structure effects.

We decided to introduce these dummy variables in the slope coefficient. Since we only obtained a significant result for the gender variable it is only for this case that we give estimation results. As shown in Table 7 we find that *female commuters are slower drivers than their male colleagues* (cf. the discussion on this subject in Sect. 3). However, it should be noted that in the present context another explanation exists for the lower speed of female drivers. It is not impossible that compared with male commuters, female drivers are more involved in multipurpose trips such as bringing children to school or shopping during the commuting trip.[7]

7. Empirical results for car poolers

The data set contains 57 useful observations on commuting distances and travel times of car poolers. A description of these distances is given in Table 8. A comparison between Tables 3 and 8 reveals that for car poolers the mean commuting distance and network time is only slightly longer than for solo drivers. For car poolers the mean values of *dc*, *dn* and *tn* are about 8% higher than for solo car drivers. One might have expected a larger difference since car

[6] An additional relevant factor concerns congestion. As already noted, the difference between reported times and network times may be partly explained by the occurrence of congestion during commuting. A possible explanatory variable would be the time of the day, but data on this are unfortunately not available.

[7] The questionnaire did not explicitly deal with multipurpose trips. Hence, it is not impossible that some respondents included additional trip purposes in their reports about commuting trips.

Table 7. Estimation results for reported travel times explained by network travel times and gender

	Independent variable is *tn*
Rounding parameter *s*	0.987 (0.0088)
Constant	7.77 (2.16)
Network travel time	0.889 (0.091)
Gender dummy * network travel time (dummy = 0 for male drivers and 1 for female drivers)	0.307 (0.108)
log-likelihood	−264.79
Gain in log-likelihood compared with specification gender dummy	1.96 (significant)

pooling is obviously not an interesting alternative for short distance trips (it is not easy to achieve sufficient compensation for the disadvantages of car pooling at short commuting distances). This is confirmed by the figures given in these tables: the first quartile of the *dc* distances for car poolers is 13.8 km, much higher than the 8.2 km of solo drivers. The fact that the means of the distributions of solo drivers and car poolers are nevertheless so near must imply therefore that also at the right hand side of the mean the distributions are different. This indeed appears to be the case: *(very) long distance trips appear to be under represented in the car pooler's distribution.* This can be observed for example for the maximum values observed. The maximum *dc* value of solo drivers observed in this data set is 75 km compared with only 48 km for car poolers.[8] Apparently at longer commuting distances the probability of a match between two commuters becomes smaller; the probability that two workers live near each other and also work near each other for these distances is small. The background of this result is that the probability of finding an acceptable job decreases strongly with distance due to the increasing commuting costs (cf. Rouwendal and Rietveld 1994). If we also take into account that the number of potential work locations at a certain distance *d* increases with *d* when jobs are uniformly distributed in space (the circumference of a circle is proportional to the radius) it follows that the probability that two workers who live in the same location will have long distance jobs near to each other is small.[9]

The relative underrepresentation of both low and high commuting distances with car-poolers is also confirmed by the coefficients of variation of the distributions. For car poolers they are clearly lower than for solo drivers.

An interesting question is to what extent **reported** driving times of car poolers are systematically higher than of solo drivers. In Table 2 we found that the mean reported travel time (*tr*) is 24% higher than the network time (*tn*). In the case of car poolers the difference is 45%. Thus car pooling appears to lead to an increase of (145/124 – 1.00)*100% = 17% of travel time.

This result for car-pooling means that where car-pooling is generally known to yield disadvantages in terms of privacy, unreliability and rigidity

[8] We are aware that the variance of a maximum value of a sample drawn from a distribution is rather high. The values found for the third quartile, confirm the pattern mentioned here.

[9] A different result would be found when the location of jobs would be strongly concentrated in one place.

Table 8. Descriptive statistics of travel times and distances of car pooling commuters in the Netherlands

	dc distance as crow flies (km)	*dn* distance according to route planner (km)	*tn* travel time according to route planner (min)	*tr* reported travel time (min)
Mean	19.4	26.9	26.4	38.4
Standard deviation	9.6	11.5	10.8	12.5
Coefficient of variation	0.50	0.43	0.41	0.32
Minimum	2.35	3.02	7.79	15
25% observation	13.8	18.2	18.6	30
Median	19.0	27.3	25.3	39[10]
75% observation	23.3	33.8	30.9	45
Maximum	48.4	55.6	60.5	60

Table 9. Correlation between reported travel time and other time and distance indicators for car poolers

	Distance *dc*	Distance *dn*	Travel time *tn*
Reported travel time *tr*	0.354	0.335	0.360

of time schedules, it also leads to longer travel times (and, possibly, travel distances).

In Table 4 we found for solo car drivers correlations between reported travel times and the other time and distance indicators to range from 0.75 to 0.80. One may expect to find lower values for car poolers; the difference is indeed very large (see Table 9). The correlations are about 0.35, implying that in an ordinary multiple regression the share of the variance that can be explained by one of these variables is only about $(0.35)^2 = 12\%$.

This low correlation makes us aware of the fact that in studies on the choice of car pooling in a transport mode in commuting it is very difficult to predict the actual transport time needed. An explanation of this result is that there is a large number of forms of car pool alternatives that will have largely varying impacts on travel times. As shown in Fig. 2 when only two car pooling commuters are distinguished, at least 10 different forms can be observed depending on the number of different work places (W1, W2) and places of residence (R1, R2), and whether or not special collection points are used (CR, CW). Depending on who of the two car poolers is driving and who is the passenger, we arrive at some 20 possible ways in which a car pooler can make the trip, each with its own implication for travel time. When more than 2 commuters would participate in a car pool we would have even more possible patterns than depicted in Fig. 3.

It is not difficult to see that the alternative forms have quite different

[10] The median value (39 min) happens to be the only observation among the car poolers that is not a multiple of 5.

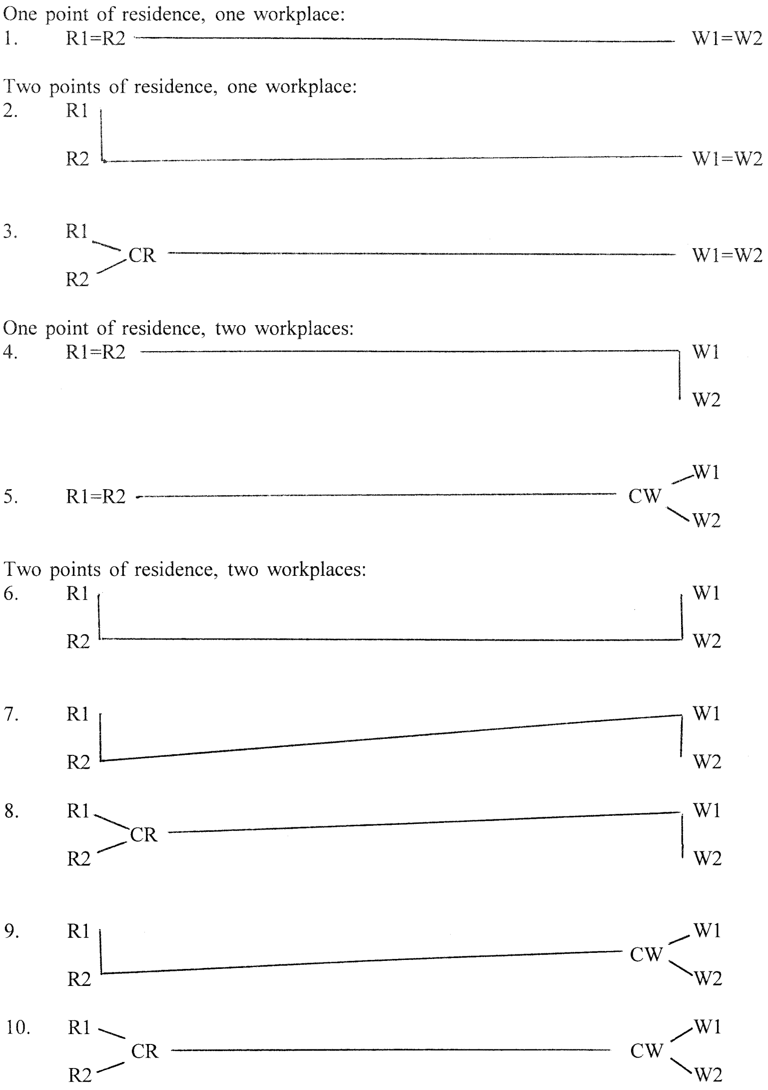

Fig. 3. Alternative car pooling structures.

implications for travel times. For example, in the case that the car poolers live in the same dwelling and have the same job location (this sometimes happens in two earner households), there is (almost) no additional travel time involved. As indicated by Teal (1987) for the USA a considerable share (some 42%) of the car-poolers are persons from the same household (types 1, 4 and 5 in Fig. 3). In that case the impact of car-pooling on the commuting time may be limited. In the remaining cases (58%) there may however be considerable effects. For example, a driver who has to make a strong detour like in case 6 will have a substantially longer travel time. In addition, in several car pool structures waiting at collection points will make actual travel time longer.

Since in our data set information is lacking about the precise form adopted, it is no surprise that the correlation between travel time and the other distance indicators is so low.[11]

8. Conclusion

The data set on which this study is based is rather small and it only refers to a part of the country. Thus one cannot claim that the results are representative for all Dutch commuters. The advantage of the data set is that it is rather rich in terms of the various distance and travel time indicators for commuters. Therefore it offers interesting opportunities for an investigation of the relationships between these indicators, which can be of use when not all of these indicators are available.

Correlations between various indicators of commuting distance (distance as the crow flies *dc*, network distance *dn* and network time *tn*) are rather high (.93–.97). The average detour factor in commuting is about 1.40, for shorter trips it is slightly higher (1.50). *We conclude that once we know the average detour factor, the distance as the crow flies dc is a quite reasonable proxy of the network distance dn.* The average detour factor found here is higher than usually mentioned in the literature (Nordbeck 1964; Warnes 1972). When sufficient data are available the average detour factor may be determined by computing it for a sub-set of pairs of points. If this data is lacking it may be estimated by judging the network structure (fine meshed, with natural barriers, etc.). Fine meshed network structures imply an average detour factor of around 1.2. When the network is broad meshed and barriers are present, it may be as high as 1.4 or 1.5.

The relationship between network time and network distance is non-linear: owing to the fact that in long distance trips the share of express ways tends to be higher, the average speed of trips increases considerably with distance.

The mean reported travel time is considerably (24%) higher than the network times predicted by route planners. The difference can be explained amongst others by non driving time components of car use, a tendency of route planners to underestimate driving times (during rush hour), and the choice of drivers for routes other than the fastest one.

Reported travel times are usually rounded in multiples of five minutes. Ignoring this fact is shown to lead to biased estimation results for shorter distances. The correlations between reported travel times *tr* and the other distance and travel time indicators (*dc*, *dn*, *tn*) are clearly lower than the those between the three mentioned indicators.

When *dc*, *dn* and *tn* are used to explain the reported travel time we arrive at non linear relationships for *dc* and *dn*, and a linear relationship for *tn*. *The distance as the crow flies dc and network distance dn are almost equivalent in their power to explain reported travel times.*

[11] Note that in our discussion about the low probability of a match between car poolers living at long distance from the work location we implicitly ignored possible forms where the distance between residential locations is long. A collection point CR at considerable distance from R1 and/or R2 would be a solution here. It seems that the simpler forms of car pooling (such as cases 1, 2 and 4) are dominant in the Netherlands.

We conclude that where actual driving times are missing in commuting research the other three indicators mentioned may be used as proxies, but that the following problems may emerge:

- *actual travel times may be considerably higher than network times tn generated by route planners*
- *the average speed of trips increases considerably with distance, implying an overestimate of travel time for long distance commuters.*

We tried to explain reported travel times by personal features of drivers. A significant result was found for gender: *other things equal, women apparently drive at lower speeds: on average they report higher travel times than men.*

In the case of car-pooling, actual travel times are much harder to explain by the other travel distance and travel time indicators. For these indicators the correlation with the actual travel time is very low (about 0.35). *On average, car pooling leads to a travel time increase of some 17% compared with solo driving*. The distribution of commuting distances of car poolers has a much smaller coefficient of variation compared with the distribution of solo driver commuting distances. In Sect. 7 explanations are given that with car poolers both lower and longer distances are 'under represented'. *The conclusion is that for a study of car-pooling in commuting the dc, dn and tn variables are very poor proxies. For car-pooling one cannot do without actual data on commuting times.*

Acknowledgement. The authors would like to thank two anonymous referees for useful comments.

References

BGC, Betrouwbaarheid reistijden (reliability of travel times), Deventer, 1996

Bovy P, Stern E (1990) Theory and models of route choice behaviour. Kluwer, Dordrecht

Jorgenson F, Polak J (1993) The effects of personal characteristics on drivers' speed selection. Journal of Transport Economics and Policy 27:237–252

Nordbeck S (1964) Computing distances in road nets, Papers and Proceedings of the Regional Science Association 12:207–220

Ommeren J van (1996) Commuting behaviour, a spatial search perspective. Vrije Universiteit, Amsterdam

Péruch P, Giraudo M, Gärling T (1989) Distance cognition by taxi drivers and the general public. Journal of Environmental Psychology 9:233–239

Rich DC (1980) Potential models in human geography. Geoabstracts, Catmog 26, Norwich, University of East Anglia

Rienstra S, Rietveld P (1996) Speed behaviour of car drivers, a statistical analysis of acceptance of changes in speed policies in the Netherlands. Transportation Research D, 1:97–110

Rouwendal J (1996) An economic analysis of fuel use per kilometre by private cars. Journal of Transport Economics and Policy 30:3–14

Rouwendal J, Rietveld P (1994) Commuting distance distributions of Dutch households. Urban Studies 31:1445–1557

Säisä J, Svensson-Gärling A, Gärling T, Lindberg E (1989) Intraurban cognitive distance; the relationship between judgements and straight-line distances, travel distances, and travel times. Geographical Analysis 18, (2):167–174

Teal RF (1989) Carpooling: who, how, why. Transportation Research A 21:203–214

Warnes AM (1972) Estimates of journey-to-work distances from census statistics. Regional Studies 6:325–326

Wee GP van (1995) Ruimtelijke en mobiliteitsreacties van werkenden op bedrijfsverplaatsingen (spatial and mobility reponses of workers to relocation of the work place). RIVM, Bilthoven

Zax JS (1991) The substitution between moves and quits. The Economic Journal 101:1510–1521