Final Report

Estimating the Value of Travel Time (VTT) using GPS data

Andrew Daly and Stephane Hess, 11 December 2018

Recommendations

Our central recommendation is that the best revealed preference (RP) VTT estimates in the current stage of technology are to be obtained from modelling mode choice using a mobile phone app (what we call type 3 data). Given the low marginal cost, we also recommend the consideration of using simple GPS data (type 1) for obtaining more accuracy in estimating the additional disutility of congestion, though this is not essential. Mode or mode-destination choice modelling based on TU data could further improve the accuracy of the estimations. These data sets would be combined and models estimated by maximising the joint likelihood. It may not be cost-effective to use mobile phone records or to make a national SC survey, unless specific VTT is required for Autonomous Vehicles, which would then be a specific survey requirement.

For modelling, we recommend that the models should remain consistent with the utility maximisation paradigm. We also recommend the investigation of a wide range of possible heterogeneities. While we do not specifically advocate the use of models incorporating a consideration/awareness stage, any future study should objectively evaluate the possible influence of the inclusion or not of such a component on model results. Inference models should be used for background variables for GPS type 1 data. Correction models should be used for network variables. Resulting VTTs should be expanded to national level using sample enumeration, which should also be used to calculate the accuracy of the VTT estimates. If modelling route choice is part of the study, then either an investigation is needed as to the biases possibly introduced by specifying choice sets a priori, or research may be required as to how recursive route choice can be modelled at national scale.

Comparisons of the resulting VTTs should be made with the recent Harbour Tunnel study and with Europe-wide meta-analysis. In making these comparisons, it should be noted that, while RP data has not been much used for VTT estimation, SC-based VTTs have not, and perhaps cannot, be validated externally, so that it is not correct to assume that these represent the correct values. SC data raises a number of important and unresolved issues also.

A decision to use RP would be based on a balanced judgement. This would need to consider the monetary budget available for the work and the elapsed time before the results would be needed.

The cost of an RP study would be more than that of a conventional SC study and it would almost certainly require more time. However, the results might be more defensible in the long term as there is indeed now a growing concern worldwide about the reliance on stated preference methods in choice modelling. Considering the components of the study:

- the cost of the survey work has a number of components, the largest of which relates to the smartphone app; if this is an app that has been translated and implemented already for Denmark, perhaps in conjunction with new TU data, then costs will not be very different from SC, albeit that higher incentives may be required due to the longer survey duration; costs will be higher if a new app is to be developed or a licence purchased for an app; standard TU data and type 1 GPS data should be available at negligible cost;
- data preparation is the step requiring the most additional work relative to conventional SC, as there may need to be map-matching work for the GPS data (or checking of map-matching work conducted by the app) and network data may need to be processed to provide times and costs;
- the modelling work may need to be extended, possibly to represent consideration sets, to model the errors in network data and to explore model specifications, although model specifications also ought to be explored extensively in SC work.

We would expect the reliability of the results and even their accuracy to be improved relative to a conventional study. In preparation for commissioning such a study, research would be needed on the likely accuracy to be obtained from given sample sizes, looking at recent RP modelling exercises where time and cost coefficients have been estimated, e.g. in The Netherlands and comparing with recent SC studies, e.g. for the Harbour Tunnel.

1. Introduction and background

The value of travel time has been said to be the most important number in transportation. Estimates of the monetary value of travel time differences between future scenarios have been used for many years by governments to appraise policy options. At the same time, VTT is applied by analysts to formulate the generalised cost of travel alternatives faced by travellers or shippers, in order to predict traffic volumes. Clearly, good estimates of VTT are necessary to obtain good decisions and good forecasts.

In Denmark, VTT for appraisal is applied using the Teresa approach and is currently derived from the DATIV project. The DATIV data was collected in 2004 and the study reported in 2007. The VTT estimated in that study have been updated repeatedly but there are increasing issues with the age of the data. Society has changed in many ways since 2004, while the updating based on income elasticities is also subject to question. Trip lengths and congestion (particularly on motorways) have also increased, effects which are only partially included in the updating of VTT, while increased use of information technology is not included at all. While models for forecasting behaviour (such as LTM and OTM) have been estimated using more recent data, there is a need to update the VTT that is used for appraisal.

Two approaches have been used for estimating VTT.

- The 'cost saving' approach (CSA) relates the VTT to the value of time in an alternative use, often working. This approach has most often been used for working time spent travelling, where the 'fully loaded' cost to the employer of an hour worked is used to obtain the VTT. However, this sort of approach was also used in early estimates of non-working VTT where figures like 30% of the wage rate were used.
- The 'willingness to pay' approach (WTP) assesses the VTT by assessing the readiness of the decision maker, usually the traveller, to pay. In the 1960s and 1970s, these assessments were based on revealed preference (RP) data, but from the 1980s onwards stated preference methods, sometimes contingent valuation (also called transfer price) but most often stated choice (SC), have been the main survey approach.

The current Danish values are based on CSA for working time and WTP for non-working time. However, for the Harbour Tunnel work, business travel time has been estimated using WTP.

The present note focusses on the WTP approach, using RP and in particular GPS data, but it is worth briefly considering the arguments concerning CSA, which has been used quite widely to obtain VTT for appraisal (and perhaps also in some cases for demand modelling).

1.1 The Cost Savings Approach (CSA)

British practice for appraisal up to quite recently was also to use CSA for working time and WTP for non-working time. However, the most recent national study has extended WTP to cover business time also, though CSA is still used for professional drivers. The reasons for the British change are not clear in the public documents. The relevant advice (WebTAG, DfT, 2017, see sections 4.1.5 and 4.2.2) simply states that WTP is used. The basis seems to have been discussions based on a report by ITS Leeds (Wardman et al., 2013), which indicates that the possibilities for estimating working time travel values are, besides WTP, CSA, which had been used before (and which is indicated as a special case of the Hensher formula), or other variants of the Hensher formula. The report avoids making any recommendation, except that different approaches should be tried and compared. However, the continuation of the CSA approach for estimating VTT for professional drivers (i.e. buses, trucks[1]) is accepted as best practice.

Mark Wardman, who was involved in these discussions with DfT, confirms that there appears to be no public justification of the switch (private communication, 2018) but advises that DfT appear to have accepted the arguments of Wardman et al. (2015) which set out a number of issues with CSA. These issues include the theoretical points that travel time changes may result in changes to leisure time as well as to working time and that such changes may also change the amount or productivity of work done during travel. Recent trends of increasing use of technology, of the proportion or employees who can use such technology and of more flexible working contracts have operated to increase the importance of these points. Moreover, CSA is not consistent with the empirical evidence of a meta-analysis presented in the paper, in that:

- it treats all time as the same, e.g. there is no premium for congested driving time, nor for walking or waiting, compared with 'in vehicle' time;
- it permits no variation by distance, except by correlation with income;
- it permits no modal variation, except as driven by income;
- it requires working VTT to increase exactly in proportion to income.

For these reasons the paper recommends a comparison of empirical results given by the Hensher equation and by WTP approaches, with the suggestion that one of these approaches might be more reasonable than CSA.

For the UK VTT study, DfT decided to estimate business VTT by collecting SC data from both employers and employees. In the event, the employee values were used, as the two sets of values did not differ greatly and the employee values seemed more reasonable; that is, the employee seems able to act as a representative of the employer. Part of the context for moving to WTP may also have been the political focus on high-speed rail (the HS2 project) and the use of travelling time for working, which is apparently omitted from CSA.

It is of course not certain where decisions on business travel are actually taken, so that even when an employer interview is conducted with a relevant person there remains uncertainty about influences on the decision by the traveller and/or other people in the employer organisation. In the present context, it is useful to note that the use of RP rather than SC data eliminates some of this difficulty of identifying the true decision-maker in travel choices, which makes it easier to justify the use of RP WTP for business VTT. However, while the focus of this note is on WTP estimation from RP data, the decision on whether this approach or CSA should be used to obtain business VTT in Denmark is not part of our brief.

1.2 The WTP approach

In WTP estimation of VTT, an important reason for the switch from RP to SC was that the RP data sets available in the 1970s and 1980s were very small (sometimes 500 or so) and, as a result, the estimates for VTT had large standard errors and thus wide confidence intervals. In contrast, the larger samples obtained from SC studies led to much smaller standard errors, even when the SC confidence limits were corrected to account for repeated observations.

Another key point that explains the move towards SC is that it is much easier to create choice situations in a hypothetical setting where a traveller faces a trade-off that involves paying more for a faster journey. In RP data, it may be difficult to sample users actually facing such trips and, for many journeys, the fastest option may not necessarily be more expensive, preventing the study of trade-offs by an analyst. In SC data on the other hand, the analyst has freedom to specify the choices that a traveller faces.

However, there remain considerable issues with SC-based estimates, largely arising from the hypothetical nature of the surveys. In common with other applications of stated preference methods

[1] Note that light goods vehicle (i.e. van) drivers are classified as business travellers, not professional drivers.

(whatever the actual format), a key issue is the lack of consequentiality. A respondent in a SC survey is not spending actual time or money, and it is reasonable to expect that the way in which a respondent accepts paying more or travelling for longer is different in reality from a survey setting. It seems a major assumption to hypothesise that the way in which real and hypothetical sensitivities differ affects time and cost sensitivities in the same way. However, this assumption is necessary if we are to claim that the relative sensitivities are in line with real world behaviour.

Additional issues with SC data, along with other interview-based data sources, are those of contact and response biases. Each survey mode has its own bias, but it would be reasonable to expect the relatively time-consuming SC surveys to bias responses towards those with lower VTT. In the Harbour Tunnel work, postcard surveys were used to try to reduce contact bias, but response bias cannot be eliminated easily. Of course, household travel surveys (such as TU) are also quite burdensome, as are surveys that involve a GPS component with a user interface. Substantial effort has gone into reducing the burden in the most widely used GPS data collection apps by improving automatic trip detection and avoiding repeated questions about the same trip on different days (e.g. commuting).

The external validity of VTT estimates from SC data has arguably not been investigated sufficiently, and analysts have focussed more on internal validity, showing that results are to some extent stable across repeated surveys. Even for analysts who are willing to assume that VTT estimates from SC data are valid, two issues in particular have dogged the estimates:

- the fact that gains in time (or money) have consistently been valued at a lower rate than losses; this effect can reasonably be claimed to be largely 'real', in the sense that people do seem to value gains and losses this way in real-world behaviour, but it is not clear whether it is partly a survey artefact, or partly or wholly a short-term effect;[2]
- the fact that small time differences (gains or losses) have consistently been valued at a lower marginal rate than large time differences; again this may be partly or wholly a survey artefact and/or a short-term effect.

For several excellent reasons, governments require a uniform VTT to apply in all circumstances. It is possible to calculate an average value for gains and losses, though this requires an assumption of how the averaging should be done, but it is not possible to eliminate the effect of the amount of the time gain or loss, so that national studies have typically been forced to make an arbitrary assumption.

Stated preference methods are increasingly criticised, whether it is contingent valuation approaches being discredited by for many purposes by McFadden & Train (2017), or the major interest in hypothetical bias in SC methods (see e.g. Harrison, 2014). For this and other reasons, including the above points about gains/losses and small time gains, SC-based VTTs are more often being called into question. Also, there remains the fundamental issue with all stated preferences, that these are not real market valuations derived from observation of actual behaviour, as classical economists would prefer.

Across different fields and for different application topics, the use of RP data is gaining renewed interest, partly due to the arrival of new types of data and data in greater volume. For these reasons, the use of RP estimation is also being reconsidered for VTT work.

1.3 Appraisal and Demand Forecasting

As has already been noted, VTT can be applied in both demand forecasting and appraisal. In principle, there should be no difference between the VTT needed in these two applications. Appraisal can be seen as integration of the demand function with respect to utility and the utility over which the demand curve is defined ought to be the same as the utility in which user benefit as used in appraisal is denominated. However, differences arise for good institutional and technical reasons.

Institutionally, governments will specify the VTT to be used in appraisal to be uniform in certain respects: across trip lengths, across modes perhaps and even across income groups. These specifications

[2] We note that the Harbour Tunnel SC considered only time gains, not losses, which means it is not adequate for estimating VTT for appraisal.

are made for reasons of equity and to make appraisals more straightforward to conduct and explain. These uniformities are of course inconsistent with the best demand modelling practice, which stresses the need for behavioural differences across travellers. The uniformity requirements do not need to affect the way in which VTT is determined, as aggregations are easily enough made when varying VTT has been estimated. Governments may also have an approach to accounting for cost that is not consistent with demand modelling and they then need to be aware of the potential issues.

The technical requirements for estimating VTT, however, can mean that a different focus may be required between demand modelling and appraisal. In the key example of route choice, we see that cost varies little across the alternatives, as they are all about the same length, so the central issue for forecasting is to represent the influence of time and congestion. Route choice forecasting models do not need a very good estimate of VTT, unless tolls are being considered. But for estimating VTT for appraisal, good estimates of sensitivity to cost are indeed required and modelling contexts have to be found in which accurate cost sensitivity estimates can be made.

In this note, we focus on the estimation of VTT for appraisal, with the secondary possibility that the values obtained could also be used for forecasting.

1.4 Structure of the note

RP data is considered in more detail in the second section of this note. There we discuss the use of traditional forms of RP data, which have been used in other transport planning contexts for several decades, such as trip diary surveys. These are compared with newer data forms, in particular including data from GPS tracking but also including mobile phone records. The advantages and limitations of these are compared. We also contrast different types of GPS data.

The use of RP data for VTT estimation is the subject of the third section of the note. The issues that arise are first to determine which choice is modelled: consideration can most obviously be given to the choice of route or mode. Other travel choices may seem to involve too many complicating factors, although destination choice is often included in large-scale models that could form a source of RP VTT estimates, and some work has looked at residential location choice; departure time choice is another possibility. The second issue is that of the choice set. For mode choice, a range of methods can be used with different levels of subjectivity that raise issues of bias or require explicit modelling. For route choice, the generation of alternatives is a considerable issue, though the use of recursive logit addresses these at the expense of modelling complexity. Then we consider the applicability of GPS and other RP data to different vehicle types, i.e. trucks and public transport as well as cars. In modelling route choice, consideration has to be given to obtaining variance in the cost variable, which can be limited in many contexts, thus suggesting that mode choice might be a useful approach. Additionally, we discuss the possibilities for obtaining VTT for autonomous vehicles.

The final section summarises the findings of the work. These give an overview of the possibilities for using GPS or other RP data for VTT estimation at a national scale in Denmark. The difficulties that would have to be overcome are listed, including the issues of expanding data that has little socio-economic content to be nationally representative.

2. Nature of RP data

In this section, we contrast new types of RP data with traditional travel surveys. We go beyond talking about GPS data alone to allow us to provide the required background but also to discuss possible alternative future data sources, such as mobile phone data. Whether or not these are to form part of the data sources in Denmark, there are lessons that have been learned when working with such data which are relevant to GPS work too. A summary of this discussion is given in Section 4.1.

2.1. Standard RP surveys

The RP data used in the earliest VTT estimations came from surveys of single trips, for which travellers reported the details of that journey.

Another source of RP data that could be considered for VTT estimation is travel surveys in which respondents report the detail of their trips, usually those made in a day, or, in some cases, over several days or weeks using travel diaries (see e.g. Axhausen et al., 2002). Such surveys are regularly conducted in several countries, to give an overall view of changing travel patterns, and are also often used for modelling travel frequency, destination and mode choice, though they are often criticised in that context due to recall issues with respondents omitting a specific selection of their trips. Several countries, in particular Sweden, have experienced problems with declining response rates in these surveys, but we do not know whether response rate is a particular problem with the Danish TU survey.

In the context of VTT work, the obvious approach with trip diary data may be to look at a single trip that is representative of the respondent's travel behaviour, for example their commute journey. A number of key issues arise:

- For many origin-destination pairs, there is a single obvious best option, and there is limited real world potential for trading between time and cost. An exception for car travel is the case of a choice between a toll road and a free road, but these are often irregular journeys, or choices faced by only a small subset of travellers. In Denmark, such journeys are very rare. For public transport, there is also not generally a choice between a cheaper but slower option and a faster but more expensive option. The uniform Danish fare system further reduces the potential for cost trading.

- Another issue is that, in the case of regular journeys, the specific route and maybe also mode choice is the result of a decision made some time (maybe a long time) ago. Trying to explain that decision with the time and cost variables for different options at the point in time where the data was collected may thus not reflect the behaviour that led to the actual choice. For instance, the current time and cost of the alternative journey may not be correctly known to the traveller.

The above issues are one of the key advantages of SC data, where an analyst can create choice situations that "force" a decision maker to make a trade-off by creating scenarios where one journey is cheaper but an alternative journey is faster. Additionally, the choice in a SC setting is more likely made on the basis of those attribute values shown to the respondent at the time of making the choice rather than some previous values. However, the SC context means that the way in which variables are defined, in particular congestion, must be in terms comprehensible to respondents. Note that this may introduce inconsistencies, as the experience of congestion in Copenhagen is different from that in Jylland.

A key point to raise in this context is the distinction between using route choice or mode choice data for VTT estimation. In the case of SC data, research has focussed very largely on the use of route choice data, partly to avoid the impact of modal preferences, but also given that in SC data, it is possible to work in a route choice setting and present meaningful trade-offs. It seems likely that for RP data, the use of mode choice may present a more suitable approach for estimation of VTT, a point we return to below.

Trip diary data is often used for general travel demand modelling and those models often contain independently estimated time and cost parameters whose ratio can be considered to be the VTT[3]. Usually, the VTT implied by these ratios is used only as validation of the travel demand models but it can also be considered as an independent estimate of VTT. Recent work in Sweden (Varela et al., 2018) has shown that the transport network data used for this modelling contains specific biases, but that these can be corrected at the same time as estimating the models. This opens the prospect of using trip diary data on its own or in conjunction with other RP or SC data for estimating VTT.

[3] Recent developments of the OTM in Denmark have not made separate time and cost parameter estimates, but have used SC-based values as constraints. We do not know what has been done in the Danish National Model (LTM) work.

2.2 Different types of GPS datasets

In this section, we discuss different possible types of GPS datasets that can be used for VTT estimation. While the type discussed in Section 2.2.1 seems to most closely represent what is currently available to VD, the type in Section 2.2.3 is the most widely used type of GPS data for travel demand modelling and is also in line with future plans for the TU survey.

2.2.1. Type 1: GPS data from loggers installed in vehicles

An approach that has been discussed is to rely on GPS data that is not collected as part of a survey but by GPS loggers installed in vehicles, as used for example by DTU in their Harbour Tunnel work. We refer to this type of data as type 1 GPS data in the remainder of this note. This is similar to the type of data used in Hess et al. (2015) for heavy goods vehicles in the UK. An advantage of this type of data compared to GPS data collected as part of a wider survey (as in Sections 2.2.2 and 2.2.3) is the reduced cost of data collection, partly as this data is made available through other sources and no new survey coding is required, and also as no incentives for survey participation are usually required.

In contrast with data linked to a wider survey, the analyst has reduced or no control over the representativeness of the sample, and reduced or no information on what are usually key drivers of travel behaviour, such as socio-demographic information and trip characteristics such as journey purpose. Additionally, these types of datasets are usually uni-modal and are limited to private cars or fleet vehicles (such as lorries, vans and taxis), thus excluding both public transport and slow modes (walking and cycling).

The DTU work to develop route choice models as part of the Harbour Tunnel study in 2017-8 used five data sets, two very large sets provided by Vejdirektoratet, one for cars and vans and the other for trucks. Developing simple (linear) models from these VD data sets, they found significant estimates for both time and cost coefficients, but in the case of trucks the estimate of the cost coefficient was significantly positive and in the case of cars/vans a rather high estimated VTT, i.e. a low cost coefficient, was found. DTU continued to develop more sophisticated models but did not arrive at convincing results for the VD data in the very limited time available. Clearly, if GPS data is to be used for future VTT estimation, more time and resources will be needed for the modelling process. Possibly, the use of log variables for time and cost, as in the Hess et al. work (2015) could be helpful. In general, it may be said that, in the absence of tolls, estimating cost sensitivity from route choice is not easy, requiring a large data set, good estimates of driving costs and probably an extensive specification search. There also needs to be variability in speed across routes. For example, the models developed from the small Borlänge data described in several papers (see e.g. Frejinger et al., 2009. or in the recursive logit work of Fosgerau et al., 2013) do not include cost or distance variables. GPS type 1 data could be expected to give good estimates of the relative values of several other components, but not of cost.

While as noted above, the original sample from a type 1 GPS survey is likely to be biased in terms of representativeness, issues with subsequent drop-out rates are likely to be very low or non-existent, largely because there is no user involvement. There is also little or no risk of missing trips in such data, except of course trips made with other vehicles or non-motorised trips.

2.2.2 Type 2: GPS data from handheld loggers, potentially combined with basic surveys

There are some examples of data collection efforts using GPS loggers not tied to a specific type of vehicle, such as in the 'tagmyday' study in Italy, and data from such studies can be and has been used for VTT estimation (Calastri et al., 2018b). These loggers are carried around by respondents during the course of the survey period, and thus capture their trips across multiple modes. This type of data collection was popular in the early 2000s but has subsequently been superseded by the approaches outlined in Section 2.2.3. Many surveys of this type also use an online portal where respondents are asked to provide additional information about their trips (see again Calastri et al., 2018b).

A key issue with this type of data is that trips are only recorded if respondents carry the logger around with them. Additionally, the fact that any online portal for additional data collection is separate from the actual GPS device increases the rate of non-completion (and therefore potential bias) of these additional survey questions.

The sample from a type 2 GPS survey is likely to be less biased in terms of representativeness than a type 1 GPS survey, because the analyst can have greater control and involvement. However, issues with drop-out rates are likely to be increased, given there is now some burden on the respondent. Additionally, there is an increased risk of the respondent not taking the logger for certain trips or turning it off, leading to bias in the reported trips.

2.2.3 Type 3: surveys enhanced with automatic GPS-based trip collection

While traditional RP surveys rely on respondent recall, there has been a shift in travel surveys towards surveys in which details on trips are captured largely automatically, generally through a smartphone app, and the respondent is only required to provide some additional detail on the trips, significantly less than the information collected in a traditional survey (purpose, travelling party, etc). For a recent discussion of such surveys, see Calastri et al. (2018a).

This type of approach has a number of key advantages.

In comparison with the more traditional surveys, the advantage is that respondent burden when it comes to trip information is reduced substantially, potentially leading to lower rates of survey attrition, and fewer missed trips. Indeed, details about the trips are captured largely automatically, and only a few simple questions need to be answered for each, while the number of these questions reduces over subsequent days as the app begins to "learn" about usual trips. Additionally, the data on the actual trips is recorded with a greater level of detail and reliability in terms of timing and routes than would be possible with respondent-provided data.

In comparison with type 1 (and some type 2) GPS data, information is collected across multiple modes, and the level of detail about the traveller and trips is at least as good as with traditional surveys. In comparison with type 2 GPS surveys, fewer trips should be missing as respondents are less likely to forget taking their phone than their tracking device. On the other hand, they can disable tracking services on their phone, which could mean some trips are not captured.

The cost of this type of survey is of course higher.

An in-depth discussion of the advantages as well as issues involved in collecting data using type 3 GPS surveys is given in Calastri et al. (2018a). This notes that while drop-out rates are of course not trivial, the biggest issue seems to be initial engagement with the survey. Once respondents have started using the app, 70% used it for the full two weeks and completed all questions. The Calastri et al (2018a) paper uses the RMove app which has been used for a number of household travel surveys in the United States[4]. Another leading app in the field, FMS[5], has similar capabilities. These are almost ready-to-use apps, which can be further customised for specific application areas, including translation to other languages.

The two apps named above (RMove and FMS) are not free, and costs vary on a case by case basis and would need to be determined by VD. They are likely to be higher than with paper based surveys. Free apps also exist, but their functionality is much reduced, and app support is often discontinued (such as with Moves[6]). Some research teams also develop their own apps, but we would not advise this in the present context[7] as this requires highly specialised skills and is likely to be more expensive and lead to a lower quality product (cf. Montini et al., 2015).

[4] https://rmove.rsginc.com/
[5] https://its.mit.edu/future-mobility-sensing
[6] https://www.moves-app.com/
[7] We are aware of some internal discussions at DTU about undertaking development of such an app in house.

Samples from type 3 GPS surveys are likely to be of a similar level of representativeness as traditional surveys. The sole additional issue is caused by travellers who do not have a smartphone. In the United States, a solution to this was to provide such travellers with a smartphone or to use a supplementary telephone survey sample. On the other hand, the novel and interactive nature of the surveys may make it easier to convince younger and more technology-versed travellers to participate. The issues with missing trips should in theory be less than those with type 2 GPS surveys, as it is less likely that the phone will be forgotten than a GPS logger that has no other use. However, with mobile phones, there is the potential of survey participants turning off location services at times to reduce batter consumption, leading to missing trips during such times.

2.3. Non-GPS sources of big data

While not the topic from the outset for the present report, it is important to highlight that various other high resolution and high volume data sources are coming on stream that have the potential, at times already demonstrated, for use in travel behaviour modelling. Key examples include smartcard data such as used in many public transport cities around the world. While such datasets have much wider reach in terms of sample than any of the approaches in the three earlier sections, they are similarly affected by limitations in terms of socio-demographic information (e.g. income) and the journeys they capture tend to be uni-modal or at best choices between different public transport options, often with limited scope for time-money trade-offs. The way that the Rejsekort system works in Denmark means that the variation of fare for a given OD (by route or by time of day) is very limited, so that trade-off possibilities in RP data involving cost have to focus on mode and/or destination choice. Nevertheless, we do not rule out using data from the Rejsekort system as a future source for VTT estimation, providing privacy concerns can be overcome, though academic research may be needed first.

An alternative to smartcard data comes in the form of mobile phone data, either GSM data or call detail record (CDR) data. The former has much higher temporal resolution as data is captured whenever the phone is turned on, but the data is not generally stored for a long time or available for modelling work. CDR data on the other hand is stored for billing purposes and often available for modelling. The temporal resolution of such data is improving given the growing use of data services, reducing the discontinuity in traces between subsequent calls. CDR data is available in very large samples often capturing a representative group of users, but like most other non-survey data types is lacking information on the actual users/travellers. The data has been used successfully for travel demand modelling (see Bwambale et al., 2018a, and the references therein). Of course, getting access to such data involves negotiation with mobile phone operators and may thus not be short-term solution in Denmark.

2.4. Data processing

A point often not covered in detail in the comparison between different possible RP sources is the amount of processing required to make the data usable for modelling analysis.

For traditional surveys, where the respondent reports a single trip or multiple trips manually, the base assumption might be that the amount of data checking that is required is limited. This is true insofar as we would expect few or no mistakes by the respondent in reporting origins and destinations (providing these concepts have been communicated clearly), but data provided on trip timing may be of low quality, especially in recall data. A step of 'data cleaning' is usually required to remove obviously illogical records.

With automatically collected data, of whatever type it might be, the issues in terms of the accuracy in terms of the timing recorded in the data are significantly reduced. The spatial accuracy however varies across types of data and even varies by the quality of the GPS receiver and the quality of the signal, which is area dependent. Even in ideal situations, the spatial and at times temporal[8]resolution of GPS data is such that map-matching work is required to allocate the GPS points to specific segments of the transport or road network (cf. Hess et al., 2015). This issue is even more severe in the case of mobile

[8] Not all GPS data sources use the same fine level of temporal disaggregation.

phone data where the resolution is poorer than with GPS (see again the discussion in Bwambale et al., 2018a).

A separate issue is that with tracking data, the start and end points of trips need to be determined. In the case of surveys using type 3 GPS data and some surveys using type 2 GPS data (those with an online smartphone interface), the user will be prompted to complete trip details that include confirming that the start and end points of a journey have been correctly identified by the *tool* (cf. Calastri et al., 2015a). A conscientious user would be expected to correct any obvious mistakes. In the case of data without a user interface, such as type 1 GPS data and some type 2 GPS data, algorithms are used to infer trip start and end points, i.e. breaking the continuous stream of data into separate trips, typically by looking at the length of stops (cf. Hess et al., 2015). These methods clearly rely on strong assumptions which will affect the resulting data and hence the models. The same applies to algorithms used for inferring modes of travel in type 2 GPS data that does not have a user interface. Ideally, models can be developed that test the effect of these assumptions, as was done in Hess et al. (2015).

In any future study using type 1 GPS data, it is essential to ensure high quality in the processing work, in terms of map-matching as well as in trip identification. While in-house processing will increase the cost, assessing and guaranteeing the quality of any pre-processing of the data done elsewhere, e.g. by the companies collecting the data, is a non-trivial task too.

A final point that deserves attention is the impact of privacy concerns on data accuracy. Respondents who have their movements tracked are rightfully concerned about privacy, and ethics requirements in publicly funded work especially will enforce strict rules. For this reason, type 1 GPS datasets especially routinely have the first and last parts of a trip (often 500 metres) removed in data processing. This of course has severe repercussions on trip identification and the ability to estimate route choice models, a point discussed in detail by Hess et al. (2015), who faced this very issue, despite dealing largely with long-distance trips. In type 2 and type 3 GPS data, the analyst generally has initial access to the full trip, from origin to destination, and while eventually some censoring for privacy issues may be enforced, this can happen after generation of the data for choice modelling analysis. Mobile phone CDR data also has some accuracy issues as a result of privacy protection measures, where for examples IDs are often scrambled on subsequent days.

2.5. Level-of-service data

An additional key step in preparing data for analysis is of course the generation of appropriate level-of-service data for both the chosen alternative and those alternatives to be included in the choice set during estimation. The specification of the choice set in a way that minimises bias also needs to be considered.

Initially, surveys relied on getting respondents to report the time and cost (and other attributes) of both the chosen option and of alternative ways of making their journey. Considerable problems were uncovered when this data was analysed more deeply, as biases of various kinds, such as rounding and 'self-justification bias' were found. For example, a respondent is likely to better remember the characteristics of the actual option that was chosen than of those that were not chosen; it is often found that the unused alternative is reported as being worse than it actually is, thus 'justifying' the respondent's choice more strongly. Despite these issues, data of this type is still occasionally used for VTT estimation, though often without great success (e.g., see Arup et al., 2015, which reports the collection of RP data but does not report on its use for VTT estimation, largely due to the poor results).

The recognition of respondents' inability to accurately report the level-of-service data for unchosen alternatives leads to a requirement to compute those. Traditionally, such data have come from network models that allow the analyst to compute travel times and costs (and other attributes) for specific routes using specific modes. Another misguided approach in this context is to use the reported (in the context of self-completion surveys) or observed (in the case of GPS surveys) level of service data for the actual chosen alternative while using times and costs from the network model for the unchosen alternatives. Take for example the case where the observed journey was heavily congested. If an analyst uses the congested journey attributes for the chosen option alongside average level of service data for the unchosen options, then the choice may appear "illogical" in the model. The assumption disregards the possibility that the congestion was not known about before starting the journey, or that congestion applied across all routes. For this reason, it is imperative and good practice to use the network level-of-service data for all alternatives – indeed using different sources for chosen and unchosen gives bias a priori, as the level of service depends on the choice and not vice versa.

It appears that the GPS data that is available gives quite good information on travel speeds and this can be used to improve the quality of network data in representing congestion etc.. Using multiple records for each link, a pattern of speed can be built up over several days of data. It must be noted that what is obtained from these measurements is a distribution of times, so that a measure of congestion can be obtained that gives the time lost relative to a free-flow journey. This is the measure that is commonly used in travel demand modelling (e.g. OTM or LTM) but is not the same as would be obtained from stated preference data (such as the Harbour Tunnel SC) where the traffic conditions experienced by the respondent are used to define congestion. Some further improvements can be obtained by looking at the GPS data in a very raw format, which will contain travel times on very short segments, often just a few metres, and an assumption can then be safely made that the average speed calculated from the distance and time reflects the actual traffic conditions on that segment.

A further issue here is reliability, but we do not know whether the number of measurements for each road segment (at each time of day) is sufficient to define a 'usual' time and delays relative to that. If there are sufficient measurements, it might be possible to build a model with an objective measurement of reliability rather than the usual subjective measurements obtained from stated preference data. However, it may be possible to build up a picture of the relationship between congestion and reliability, i.e. that the variation in time for a link is related to the average time lost (relative to free-flow).

A further complication relates to the composition of the choice set. While we return to this issue in more detail in the following section, it should already be noted that respondent-reported consideration or availability sets are subject to the same biases as respondent-reported level-of-service data, and that at best, such data should be used in a probabilistic manner (Calastri et al., 2018b).

2.6. Trip Costs

An important issue is what costs need to be associated with a trip or tour. We may distinguish four different specifications of cost per kilometre:

- *average cost*: this is the total cost of owning and driving the vehicle divided by the total number of kilometres driven, for example on an annual basis. These costs are usually available from motoring organisations and in some countries form the basis for tax-deductible allowances for driving a personal car on business.
- *marginal cost*: this is the addition to the total cost brought about by driving an extra kilometre. Again, calculations can be made based on technical issues of vehicle performance and the car fleet owned. These costs are of course much lower than average costs, as the cost of ownership and insurance etc. depend only to a limited extent on the distance driven.
- *perceived cost*: this is the impression of cost by the owner or driver on the basis of which decisions are made about driving. We cannot observe these costs, so that the only information we have about them comes from models.[9] Sometimes it is assumed that perceived cost is equal to fuel cost, but the evidence for this is very tenuous[10]; alternatively, it is sometimes assumed that perceived costs are zero, but this contradicts the results from models.
- *reported cost* is yet another measure of cost, but one that is not directly useful in this context, as drivers respond with all kinds of biases, roundings and approximations, including zero. However, reported cost is needed in SC experiments, as it is necessary to establish a basis for the experiments which is credible to the respondent.

[9] And different uniform specifications of cost per kilometre will give equally good models, but yield different VTT, so that well-founded values are needed.

[10] In the UK, the basis for this assumption seems to be a PhD thesis completed in 1966, which itself had very limited evidence and which has never been reviewed seriously.

All of these costs are functions of vehicle size and efficiency, fuel type and price, annual kilometrage and driving style, including speed. Insurance depends on the driver's age and vehicle value. Moreover, some of these characteristics are the result of choices that may depend on the trip pattern, while in other cases the trip pattern will depend on driving costs. In any case, forecasting models are not able to deal with these details, though they can give trip purpose, trip length and average road speed. National shifts in the distribution of vehicle efficiency and fuel types in the fleet can also be forecast.

In choosing which cost specification to use in the estimation of VTT by modelling WTP, it seems that average costs should be used for business travel, as this is likely to be the cost incurred by the business and which is compared against alternative ways of making the journey (public transport, taxi, etc.). For leisure travel, we recommend the use of marginal costs. The use of perceived costs appears attractive, but these would need to be set against perceived times and we have no way of knowing either of these, particularly in the context of GPS data. Moreover, marginal costs are what is actually incurred by the driver and their use then makes the appraisal framework consistent, in the sense that costs and times actually incurred by car users are compared with the actual investment and environmental costs etc. incurred by society. In Denmark, the use of average costs in the appraisal procedure introduces inconsistency, because it seems that travellers' behaviour relates better to a lower per-kilometre cost, such as is given by marginal cost.

2.7. Joint use of different data sources

The limitations of the type of GPS data (type 1) currently available to VD, with respect to socio-economic and detailed trip data, as well as its relative weakness in estimating cost sensitivity, suggest that it may be interesting to consider modelling VTT with a range of data sources. We could consider the existing TU and SC data, but also new interview data to supplement the existing GPS information, where it would then be desirable that such new interview data uses a smartphone app which would mean that the level of accuracy is comparable with the existing GPS data.

For car users, interview data would be used to give purpose and income distinctions, driver/passenger differences and possibly improved estimations with respect to vehicle type, e.g. for vans.

The greatest benefit in additional data collection would arise for public transport users. Indeed, the current GPS data would not allow us to understand the behaviour of public transport users, or indeed the choice between car and public transport. In other cities/countries, this issue has been addressed by conducting smartphone surveys with a GPS component (i.e. type 3 GPS data), meaning that both car and public transport trips are captured. However, data of this type would require some effort to set up and the cooperation of travellers to conduct, raising a possibility of response bias. Moreover, Danish fare systems appear to give little or no cost variation by route, so that time-cost trading does not occur. These features suggest that mode or mode-destination choice may be the best way to determine public transport VTT; these choices could be modelled using TU data, possibly with the proposed new smartphone component. Alternatively or additionally, SC data could be collected from public transport users, either on notional route choice within the public transport system or on mode choice. Public transport and car VTT derived in this way could be related to car VTT derived from GPS data.

VTT for other modes, such as cycling and walking, could be derived analogously. However, we advise against the modelling of VTT for walk and cycle users by SC including the addition of notional costs, as these have not worked well in previous studies (the UK study conducted such work but it was not included in the final results published in Arup et al., 2015).

In general, we would advise that the decision between continuing to rely on route choice (as used in SC studies for VTT), or moving to mode choice, needs careful consideration. In a real-life setting, such as captured by type 1-3 GPS data, or mobile phone GSM/CDR data, the within-mode trade-offs against cost are very limited, and the scope for studying VTT on the basis of mode choice might be more promising. However, within-mode trade-offs of different sorts of time, i.e. congested vs. free-flow time for car journeys or walking and waiting time vs. 'in vehicle' time for public transport, offer more potential.

A related issue is that it is desirable to make VTT estimates for travel in autonomous vehicles (AV). Of course, it is not possible to make RP estimates of these values, as AV are not yet available for general use. Assumptions could be made that using an AV was 'like' being a taxi user, car passenger or perhaps a bus passenger, although these modes are currently used by specific groups of travellers with very specific values. Recent publications suggest that the impact of AV could be to bring about quite extensive changes in behaviour and that the definition of VTT may need to be revised.[11]

An alternative approach would be to obtain VTT values for AV using SC methods. The validity of this approach would depend on our ability to describe travel in an AV in an objective way, which would be difficult given the positive and negative publicity about these vehicles. If VTT for AV is sufficiently important, an SC approach is recommended, with comparison to values in other modes serving as a validation and backup.

The chief advantage as well as the chief difficulty of using SC for AVs is that travel in such vehicles will be a new experience. Scope for using the travel time for other activities than driving depends on the extent to which it may be necessary, or seen as necessary, to supervise the automatic control of the vehicle, e.g. for safety or route finding. To conduct an SC exercise it would be necessary to determine in the scenario to be studied what level of involvement the traveller is to have (different scenarios may be appropriate), then to present this to the survey respondents in a way that is sufficiently persuasive to overcome their existing prejudices resulting from media coverage of AVs. Ride comfort would also have to be specified and described.

3. **Estimation procedures**

The previous section has contrasted the different types of RP data that might be used for VTT estimation, with a focus on various sources of GPS data. In this section, we look in turn at the implications in terms of model development and estimation when working with GPS data. A summary of this discussion is given in section 4.2.

3.1. Overall model framework

Whatever form the GPS data takes, it will contain information on individual trips. Each of these trips will represent one observation in the data, and the aim of the modelling analysis is to represent mathematically the choice of that specific trip from a set of possible alternative ways of making the same journey. This thus represents a discrete choice process, choosing a single alternative amongst a finite set of possible options.[12]

While much interest has been generated into alternative model paradigms for discrete choice, often with the goal of increasing "behavioural realism", the aim of the present analysis is to use the parameters estimated from the model to compute value of time measures and other monetary indicators. These constructs rely on micro-economic theory, and for arguments well rehearsed elsewhere (e.g. Hess et al., 2018), the models used for VTT estimation should adhere to the random utility maximisation (RUM) paradigm. As also discussed by Hess et al. (2018), many behavioural factors that are at first sight not compatible with RUM can be accommodated successfully in a RUM model. Of course, there are limits to what can be achieved while maintaining complete theoretical consistency with RUM, but before a model is accepted that is potentially inconsistent with RUM, a rigorous demonstration needs to be given that a proper VTT can be calculated from it.

[11] See the special issue on this subject in Transportation, Vol. 45, Issue 6, Dec. 2018, which contains a number of papers giving ideas about behavioural changes.
[12] 'Recursive' models allow loops in the network and so can accommodate a choice set that is in principle infinite.

Just as in studies estimating VTT from conventional RP or SC data sources, a number of decisions are required by the analyst in terms of model structure (e.g. logit vs mixed logit), functional form (e.g. additive vs multiplicative error terms, preference space vs WTP space), utility specification (e.g. in terms of socio-demographic effects) and assumptions about random heterogeneity (e.g. distributions, correlation, etc). For many of these factors, there is little difference when working with GPS data, and we will therefore not revisit each one of these points below. We instead focus on key considerations that apply when using GPS data for VTT estimation.

Aside from a detailed specification search in terms of socio-demographic effects, our recommendation in general would be that any future study makes use of advanced mixed logit models to capture heterogeneity in VTT, with flexible distributional assumptions including deviations from standard parametric distributions (cf. Fosgerau & Mabit, 2013). The work should also explicitly test for the most appropriate error structure to use (cf. Fosgerau & Bierlaire, 2009), and with the aim being the estimation of VTT and other WTP measures, the work should most likely rely on estimation in WTP space rather than preference space (cf. Train & Weeks., 2005).

3.2. Decision to model

The estimation of VTT using discrete choice models does not make any a priori assumptions about the type of decision that is modelled, and only requires that the choice process studied allows the model to estimate the relative sensitivities to time and cost components. This is possible with both regular decisions such as route choice, mode choice and destination choice, as well as longer term choices such as residential location choices.

In the context of SC data, the focus in VTT work has been on looking at route choice[13], i.e. facing travellers with a number (generally 2-3) of mutually exclusive ways of making the same journey using the same mode but with different routes. While the focus on route choice is not without complications (creating a hypothetical setting in which a faster option is more expensive is difficult especially for car travel when not involving tolls), a number of reasonable arguments have motivated the general use of route choice:

- mode choice decisions may, especially in a hypothetical setting, be overly influenced by modal preferences, rather than time and cost differences. Because of the nature of the modes, e.g. buses are generally slower than trains, specific mode preferences like the image of a bus can be confounded with the performance in terms of time. The context in which time is spent is different between modes, e.g. comparing train time with car time, so that the direct comparability of the alternatives is reduced. In addition, mode choice opens up uncertainties in relation to availabilities of different modes and consideration of each mode;
- destination choice requires the elicitation of large numbers of potential destinations, with many relevant characteristics, and reducing these to small numbers is not possible as easily as with route choice; and
- long term decisions are difficult to represent in a hypothetical setting while retaining realism (although used in the German study, c.f. Axhausen et al., 2015).

While the initial view might be that route choice has served us well for VTT estimation on SC data, we would caution against simply making the assumption that it also presents the only and most appropriate way of estimating VTT from GPS data. A key reason for the move away from RP data for VTT estimation was the lack of meaningful trade-offs, i.e. where faster options incur a higher cost. This issue is in fact endemic to the majority of within-mode choices of time against cost in a real-world setting, where lack of variation and high correlation are common, especially so for car travel (where a longer route will cost more) and in the Danish context also for public transport, while it is very difficult to attribute costs to alternative walk or cycle routes. This thus creates a potentially very significant issue for working with type 1 GPS data for VTT estimation, as seen in the DTU analyses discussed above, although it has been shown that this is possible (at least for trucks, see Hess et al., 2015). Within-mode

[13] Some studies could be seen as asking for a preference between abstract combinations of time and cost but these are not very different from route choice, which seems to be the presentation most often used.

route choice involving different dimensions of time, e.g. congested and free-flow time, should work well with all types of GPS data, however.

With type 1 GPS data that is specific to one mode, the study of mode choice is of course not possible. Studying destination choice would potentially be a possibility, but is fraught with issues as no information is available on the travellers. One possibility other than route choice when working with type 1 GPS data is to look at departure time choice, which has been successfully done with CDR data, which has many of the same characteristics as type 1 GPS data (Bwambale et al., 2018b). Again, trading against cost would be difficult because of the limited variation, but trading in different dimensions of time would be possible.

Type 2 and especially type 3 GPS data open up the possibility of not just looking at route choice but also mode choice, while the modelling of other dimensions, such as departure time choice and destination choice, are facilitated by additional information on trip purpose and, in the case of type 3 GPS data, scheduling of activities and constraints.

The view seems to have established itself in the VTT community that mode choice is less suitable for VTT analysis as modal preferences may dominate the choices. However, with suitably rich data and appropriate model specifications, there is no reason why reliable VTT estimates could not be produced from mode choice models, as seen for example in Calastri et al. (2018b). Additionally, this opens up the possibility of different VTT measures by mode for the same traveller (rather than differences in VTT across users of different modes). And of course mode choice opens up greater possibilities for meaningful time-money trade-offs in RP data than route choice alone.

An important possibility here is to use trip diary data such as TU or GPS assisted smartphone surveys in a large-scale model to estimate mode choice or mode and destination choice. Models of this type estimate time and cost coefficients with quite high accuracy, given the large data sets often available (such as TU). It seems reasonable to interpret the ratios of these coefficients as estimating VTT. Recent research in Sweden (Varela et al., 2018) has shown that the network-based time and cost variables that are used in large-scale modelling can be corrected to improve the models and reduce bias in the estimates. Further work would be needed to turn the models into VTT estimations that could compete with SC and other RP-based estimations, but the approach is highly promising at this stage.

3.3. Choice set composition: behavioural and practical

In SC data, the number of alternatives presented to each respondent is controlled by the analyst. This number is generally small, especially in the case of route choice settings, keeping computational costs low for model estimation[14]. In SC scenarios focussing on mode choice, the scenarios are often customised in such a way that only those modes that are available and reasonable for the specific type of trip are shown.

In GPS data, the issue becomes more acute. The number of alternatives available becomes much larger in a real-world setting, especially so with route[15] and destination choice.

We start our discussion of this problem by looking at destination choice to illustrate the issue, before turning to route choice where further complexities arise. For any given trip, the number of possible detailed destinations to choose from is so large that model estimation can become computationally intractable[16]. Additionally, with a very large number of possible destinations, the development of

[14] Although the main argument in SC is on reducing respondent burden and practical issues in terms of displaying large numbers of alternatives on a screen

[15] Particularly for car travel, maybe less so for public transport.

[16] Substantial headway has been made in recent years with software allowing for multi-threading and cloud computing, which offers opportunities for taking advantage of extremely powerful hardware at low cost, but this is not sufficient to make estimation on the full choice set practical or possible if working at the level of individual

appropriate level of service data becomes non-trivial. Analysts thus often estimate models on reduced choice sets, containing the chosen alternative and a subset of the unchosen alternatives. This process, commonly referred to as sampling of alternatives, reduces the full set of alternatives not for behavioural reasons, but to reduce the computational burden of model estimation. This process is fraught with difficulty. Firstly, the analyst needs to decide how many destinations to include in the choice set, where the richness of the data and hence the quality of the estimates (e.g. the accuracy and reliability) likely increases with the choice set size, but at the expense of increasing computational burden. Secondly, a process needs to be used to determine which alternatives to include in the choice set. A purely random process runs the risk of creating choice sets in which the chosen alternative is nearly dominant, reducing the quality of the data for estimation. Any more "intelligent" process however needs to use a priori information about the choice process. Except with the most simplistic assumptions about the error structure of the model and with the simplest sampling procedures (random rather than linked to behaviour), this sampling of alternatives will lead to bias in model estimates. While correction approaches exist (see detailed discussions in Guevara & Ben-Akiva, 2013), these are not trivial to apply.

With destination choice, an analyst can relatively easily enumerate all the possible alternatives, where this is simply a function of the level of spatial disaggregation (e.g. zone vs parcel) as well as the geographic scale (within the city, regional, national), and then use a subset of them through an appropriate sampling procedure. Route choice on the other hand is substantially more difficult. Indeed, especially with car travel, the number of possible unique routes between an origin and a destination, even after eliminating loops, is very large. While solutions exist for finding the shortest path (cf. Dijkstra, 1959), the situation becomes more complicated if we want to create a set of *K* 'significantly different' paths to include in the choice set. A review of techniques is included in Hess et al. (2015) and the references therein. The process is again computationally difficult, especially for large *K* and for networks with many vertices. Any study using route choice and estimating models on a subset of alternatives should thus test the impact on model estimates of the decisions about *K* as well as the impact of the algorithm used to construct the set of *K* alternatives for each trip. The impact likely reduces with larger *K*, but this of course leads to increases in computational cost both in choice set generation and in model estimation. Additionally, the choice set generation process is likely to be improved by not just focussing on physical distance in a shortest path algorithm but on a generalised cost measure that incorporates other route characteristics (e.g. Antonisse et al., 1988). This however again requires prior assumptions about the specification of such a generalised cost function. Either way, the use of a subset of alternatives is unavoidable with a model for route choice where the alternatives are individual routes between the origin and destination. An alternative approach, which we look at in the following section, is to instead model the choice as a sequence of decisions at each node along a route.

The use of a reduced set of alternatives as discussed above is largely based on computational considerations. We still assume that the "real" choice set used contains all the alternatives and we approximate the choice process by using a subset. However, there is also a substantial literature that looks at the possibility that respondents themselves may focus on only a subset of alternatives. This is most easily explained on the basis of a two-stage process, where individuals first determine which alternatives to consider and then making a compensatory choice amongst the remaining options. In the case of mode choice, the argument is that certain modes may be systematically excluded from the choice set, while, across settings (mode, route, etc), there is also an argument that respondents may exclude alternatives if they perform particularly badly on one characteristic, no matter how well they perform elsewhere. As an example, if a given route exceeds a certain time threshold, it will not be considered, not matter how cheap it might be. Different approaches have been used in the literature to accommodate such behaviour. The majority of work has looked at the probabilistic inclusion of alternatives in the choice set. This uses a latent class model, where each class has a different choice set, but uses the same parameters in the utility function. This type of structure was first put forward by Manski (1977), who does not explicitly link the existence of different probabilistic choice sets to consideration by the respondent but to an undefined process which generates the choice sets. In using this type of model for

properties. When working at the zone level, full choice set is estimation is possible with powerful software and hardware.

consideration, we recognise that the actual inclusion of an alternative in the choice set is not observed, and is thus treated as latent, making the inclusion of a given alternative in the choice probabilistic (see e.g. Swait, 2001). This can lead to substantial increases in computational complexity of the models, as the choice model now becomes a latent class structure, with different consideration sets in different classes, where, in the full Manski model, there are 2^K–1 classes when working with K alternatives. In other work (see e.g. Cascetta and Papola, 2011), analysts have avoided the use of such a latent class model, instead including discounting factors in utilities to approximate in a utility maximising way the non-compensatory effect of an alternative exceeding a given threshold, for example. The use of such a "consideration" component in models is not uncontroversial, and it is clearly possible that it takes away from the main choice component of the model and potentially biases those results.

An additional dimension, and one which is more reasonable to defend, is the possibility that some travellers may in fact not be "aware" of all alternatives. A standard specification of a model might assume that all "reasonable" alternatives are known to and considered by a traveller and thus included in the choice set. What defines "reasonable" is already in the eye of the beholder and any poor assumptions by the analyst, i.e. excluding some alternatives, may lead to bias in the VTT. Conversely, assuming that the full set of alternatives is included in the choice set by the traveller may also lead to bias in the VTT if the traveller in fact is only aware of or only considers a reduced set of the alternatives.

The possibility of bias can be most easily understood by thinking about the independence from irrelevant alternatives (IIA) assumptions. If this assumption were to hold, parameters would be unbiased, independently of which of the unchosen alternatives are included in the choice set, provided this is done with uniform probability. However, empirical evidence clearly suggests that IIA is often violated in both route and mode choice, and there is then no guarantee that bias is avoided. While modelling choice set composition in SC data is already a non-trivial task, when we turn to type 1 GPS data, we know little or nothing about the travellers and even in type 2 data little about the circumstances in which they make their trips, making the process even more difficult. A further component relates to unobserved availability of options, a point we return to in Section 3.5.

An argument can of course be made that a suitably flexible specification of a RUM model with full awareness and consideration of alternatives will explain the choice data just as well as a model that incorporates probabilistic choice set composition or the role of thresholds. If however thresholds (for example) play a role for travel time, then the estimation of a marginal utility coefficient over the entire range of observed travel times may be affected by the presence of these thresholds. For example, if travel time sensitivity is linear at a rate of β_t between 0 and some thresholds T_c, but no alternatives are considered beyond T_c, then the estimation of a coefficient over the entire range from 0 to the highest observed travel time will lead to a more negative estimate for β_t. It is however also possible to see major disadvantages in the inclusion of a consideration stage, whether probabilistic or via thresholds. Indeed, the existence as well as role/extent of consideration and thresholds is not observed by the analyst, and the true choice process may in fact be compensatory, or at least compensatory for some people. The incorporation of a consideration component in our models however moves us away from a compensatory decision process and also creates the issue that the resulting VTT estimates would only relate to the compensatory part of the model, i.e. within the range where the model hypothesises that alternatives are considered.

In this note, we do not advocate that a future study should include awareness and consideration in the models. We do not take a position on whether a "choice set" or "consideration set" actually exists for each traveller and whether this is different from the "universal" choice set. We simply highlight that these issues have received a substantial amount of attention in the literature, and that any future study should evaluate the benefits as well as pitfalls of either using the full set of alternatives or a reduced set.

In the context of GPS data, the topic of the choice set to include in the modelling thus has two distinct components.

- Is it reasonable to assume that all available alternatives are known to and considered by the travellers when making their choices? If not, how would an incorrect assumption of full consideration bias VTT, and how can consideration be incorporated in the models in a realistic and computationally feasible manner? Conversely, is it reasonable to assume that only a subset of alternatives is considered, and what is the impact on results if this assumption is incorrect, or if the subset used in the models does not correspond to that actually considered by decision makers?
- Is the number of alternatives that are to be included in the final choice set too large to enable estimation of the model within reasonable time and with standard hardware, or to even generate the full set of alternatives? If not, what sampling of alternatives is reasonable and what corrections for the resulting bias are needed? What is the impact on results of the choice set generation process used especially for route choice.

With mode choice, the computational issues discussed above do not arise. With destination choice, the requirements for sampling depends on the level of spatial disaggregation used. With route choice on the other hand, the question arises whether sampling of alternatives for computational reasons can be avoided by reframing the choice process, a point we look at in the following section.

In summary:

- While we do not take a position on the use of awareness or consideration components in route choice models, any study doing so needs to test the impact of these components on model results, in particular the impact on the VTT estimates.
- The use of sampling of alternatives for computational reasons is likely required in route choice models (except if using the approach discussed in Section 3.5) as well as in destination choice with fine spatial resolution. It is essential that studies using sampling carefully consider the corrections required to avoid bias (cf. Guevara & Ben-Akiva, 2013), which become more difficult with increasing model flexibility, which is relevant given the strong reliance on mixed logit in VTT work. In addition, the studies should test the impact on VTT results of using different sampling approaches, both in terms of the algorithms used to sample the alternatives as well as in terms of the number of alternatives included in the sampling (see for example Hess et al., 2015).

3.4. Model structure: choice process and overlap of alternatives

In the context of mode choice, but also destination and departure time choice, the choice can be seen as a single decision taken at one time. This implies that for modelling, the choice is modelled as a discrete choice from a subset of mutually exclusive alternatives.

For now, we retain the assumption that the choice is made once for a journey, at the outset, or at least that we can use this as a reasonable approximation to the choice process; in section 3.5 we discuss models avoiding this assumption, i.e. models that do not depend on an assumption about when the decision is made or if the decision is in fact a sequence of decisions. Choice is made from either the full set of alternatives, or from a subset as a result of either consideration of alternatives and/or sampling for computational reasons. Either way, the number of alternatives is likely to remain much larger in each choice set than would be the case for most SC studies. A key issue to consider when working with GPS data is then the similarity between the alternatives in a given choice set. This takes two quite distinct dimensions.

The first point relates to alternatives sharing common unobserved traits, leading to correlation between the error terms for those alternatives. This can be seen to apply especially in the case of mode choice, where for example rail and bus may be closer substitutes for each other than car and bus. This correlation can be most easily understood by noting for example that some travellers will have a preference for public transport and others for private transport. This unobserved heterogeneity affects the error terms for the different alternatives, where this impact is shared by bus and rail. Accounting for such correlation in the error terms is commonly done using models with more flexible error structures than multinomial logit, such as nested or cross-nested logit, or ordered generalised extreme value for departure time choice. The correlation can also be captured using an error components specification of a mixed logit model, as done for departure time choice by Paag et al. (2001), for example. Accounting for unobserved similarities between alternatives adds some computational complexity, but of course improves the models and reduces bias in the resulting WTP measures such as VTT. Crucially, the use of either a nesting structure or an error components mixed logit model retains the properties of the RUM framework, allowing us to compute VTT consistently with the economic framework.

The second point relates to the somewhat different situation when it comes to route choice. Each route between an origin and a destination is made up of individual links or segments, and the different possible routes between this origin and destination have different degrees of physical overlap with each other. For a given route j, some of the other routes will not share any links with j, while others will use the same links as j for the majority of the route. This overlap between routes can be observed by the analyst.

However, using a nesting structure for route choice is not as straightforward or in fact as possible as one would like. Firstly, the use of a nested logit structure is not feasible as it would only allow for single nest membership and would thus not be able to capture the impact of the different overlap between routes. Various approaches have therefore been used in the literature to capture such correlations, using either more advanced nesting structures such as cross-nested logit, or probit or error components logit. For a detailed review, see Prato (2009). These models all add substantial complexity, further complicating the analysis on data already affected by large choice sets. An innovative solution was put forward by Frejinger & Bierlaire (2007), where the correlation between routes is driven through their use of specific subnetworks of the overall network, such as key routes.

An alternative stream of research has taken a more direct approach, by not using correlated error terms, but by introducing correction terms into the deterministic utility function to approximate that correlation. A priori, it is reasonable to assume that a route that has more overlap is penalised for that overlap and thus attracts a reduced probability. The question arises as to how to define these terms and whether the resulting model remains consistent with RUM and thus suitable for VTT estimation. Two broad groups of models exist. The commonly factor (or C-Logit) models penalise an alternative as a function of the length of the route shared with other routes (see Cascetta et al, 1996, or the later references in Prato, 2009). The path size (PS-Logit) models on the other hand work out what part of a given route is "unique" and favours alternatives that are have more unique links (see Ben-Akiva & Bierlaire, 1999, or the later references in Prato, 2009).

We now return to the question of consistency with RUM, which is essential for VTT estimation. A key characteristic of RUM models is that the utility of an alternative depends only on the attributes of that alternative and is not affected by the existence of other alternatives or their attributes. The impact of these other alternatives on the probabilities is captured by comparison of utilities across alternatives and also through the error structure of the model. In both C-Logit and PS-Logit, the utility of a given alternative j is directly affected by the existence of other alternatives, and by the specific routes used by those alternatives, through the overlap these create. In the simplest specifications, the impact of alternative l on the commonality factor and the path size for alternative j is governed only by what links route l shares with route j. In more advanced specifications, the length of those links also enters the calculations. The preference for route j over route k can therefore be influenced by the existence of l, (in the more advanced specifications, also by the length[17] of l), thus undermining the utility basis of the model (which requires that the utility for alternative j depends only on the attributes of alternative j). A further complication arises if we also use sampling of alternatives, whether for choice set generation or for computational reasons. If the inclusion of an alternative in the choice set is a function of its attributes, e.g. because only a fixed number of the best routes are considered, then the path size and commonality factor also depend on those attributes, given that they are a function of the choice set; this could raise

[17] In this context, 'length' is usually simply geographic length, which does not raise new issues with respect to RUM, but if length is defined more abstractly to include time or cost (as in some work) then even more serious issues of consistency with RUM can arise.

issues of potential bias in the model, which could affect all of the parameters, including the VTT. This is not to say that C-Logit and PS-Logit do not give a good representation of route choice behaviour, rather just to say that it is not a priori clear that they can always be interpreted as fully utility-based models. The importance of capturing overlap between routes is however also clear, precluding the use of simpler model structures, and the use of models capturing the overlap in the error structure is not straightforward.

Let us now consider the calculation of VTT from such models. The calculation of VTT is based on the notion that a change in one attribute (time) for a given alternative can cancel out the change in another attribute (cost) and keep the utility and hence probability of the alternative the same all else being equal (i.e. no changes to other alternatives). The VTT is then given by the ratio of partial derivatives of the utility function in relation to the time and cost attributes[18]. In a typical linear-in-attributes utility specification, these partial derivatives are constant, meaning that the VTT is given by a ratio of the time and cost sensitivities, where these may of course vary across individual travellers. If time and cost enter the utility function in a non-linear way, the VTT also becomes a function of the time and cost of the trip. However, the way in which for example increases in cost can be compensated by reductions in time for the same alternative remains independent of the characteristics of the choice set, i.e. the existence or attributes of other alternatives. If these other alternatives stay the same, then there exists a single change in time that can compensate for a change in cost and keep the probability of the alternative constant. Of course, in a model such as random regret minimisation (RRM), it is clear that this is no longer the case, because regret is calculated with respect to all the other alternatives, and the relative impact of time and cost changes now becomes a function of the existence and attributes of all alternatives in the choice set. The amount of time reduction required for a given alternative to compensate for a cost increase for that same alternative and keep its probability constant now depends on the existence and attributes of other alternatives. The calculation of a VTT measure then becomes context dependent and can no longer be seen as suitable for welfare analysis (cf. Hess et al., 2018). The inclusion of commonality factors or path size into random utility models means that the utility of a given route i is now a function of the existence of other routes. The impact of other routes on the utility of alternative i however only depends on the existence of these routes and the links that are shared, as well as the length of those links depending on the specification used. While this could be seen as raising possible doubts about the theoretical consistency with RUM, as the utility of i now depends on the specific routing of alternative j, this impact is fixed and does not change with for example changes in time or cost (unless these are included in the commonality or path size variables). Furthermore, the partial derivatives of the utility of alternative i in relation to time and cost are not affected by the presence of these terms in the utility function (although the marginal utility parameters may of course differ from a simpler model) and the VTT is thus not context dependent.

The key point in terms of guidance for using models of the type above is then that:

- Models incorporating approaches to accommodate the overlap between routes need to carefully consider the implications this might have in terms of consistency with RUM and should test the impact of these additional model components on the VTT estimates.

Another contribution to mention in this context is the work of Frejinger et al. (2009). In their work, the underlying assumption is that all possible routes are included in the choice set of the decision maker and that importance sampling from these routes is used for the actual econometric model, thus estimating on a subset of the routes, but with an explicit correction of the utilities to account for the sampling. They also propose a revised version of the path size correction which again is derived from the sampling protocol. In practice, it is not clear how easily applicable the Frejinger et al. (2009) approach is, as their empirical work is based on a synthetic network with only one-way roads and no loops. With a larger network, the ability to calculate path selection probabilities is no longer obvious.

[18] In some non-linear models it is necessary to consider finite changes to time and cost that keep the utility of an alternative unchanged – see (e.g.) Hess et al. (2017).

3.5. Recursive choice models

For route choice, we have already indicated that the number of possible routes leading from an origin to a destination is potentially very large, especially for private transport, and generally too large for all routes to be included in a modelled choice set. In reality, the question arises whether this is a realistic way for travellers to make their choices, or whether travellers make decisions during their journey. This would be in line with step selection models commonly used in ecology (Fortin et al., 2005) and would imply that a traveller makes a decision on which way to turn at every junction in a network. In reality, a mixture of the two is likely to apply, with some preplanning of an overall route and some decisions along the way, at specific junctions.

We have already discussed above how the representation of the choice of a route as a one-off decision at the start of a journey leads to computational issues given that the resulting choice set of possible routes is very large even in small networks. The use of sampling of alternatives becomes essentially unavoidable, with all the problems this entails. An alternative approach is to model the choice as a sequence of decisions at the level of individual nodes in a network, where the resulting choice set for each decision is thus much smaller, containing only the possible links leading out of that node. This is the premise of the recursive logit (RL) model put forward by Fosgerau et al. (2013), which substantially improves on the model proposed by Dial (1971). While the main motivation of the development of RL seems to have been to allow an analyst to avoid sampling of alternatives, it also moves us behaviourally towards a step selection model, especially when incorporating discount factors, as outlined towards the end of this section. Our discussion of these models is based only on review of the literature, as we have no personal experience of working with them.

In the RL model, the probability of a traveller choosing a given route (as observed in the data) is given by a sequence of probabilities of decisions at the nodes along that route. At each node (e.g. a junction in a road), the decision maker chooses amongst the available links leading from that node, where this includes the previous link too, thus allowing for u-turns[19]. The decision at each node is then expressed as a multinomial choice, modelled by a multinomial logit (MNL) model in the case of the RL model. The choice set at each node is limited to the set of links that originate from that node, which, in any realistic setting, will be sufficiently small so as not to cause computational issues.

The key issue in operationalising the RL, i.e. calculating the probabilities, relates to the definition of the utility for a given link in the choice between all the possible links leading out of a node. Say we are looking at the choice between the different links out of a node v_0. In the RL model, the utility for a link l is made up of two components, namely the "instantaneous" cost, which is the cost of that link l itself, e.g. the time to the next node, say v_1, and the maximum expected utility across all possible routes leading to the destination from node v_1, i.e. the best possible route to the destination after taking link l to reach v_1 from v_0. This is where the recursive nature of the model arises, given that the expected utility of the remainder of the route after taking the specific link l is again given by the maximum across all the links that lead on from v_1, each time of their own instantaneous utility summed with the expected downstream utility. At each step in this recursive process, an error is added to the utility. This expected utility for a given link is given by the Bellman equation (Bellman, 1957) which needs to be solved to estimate the RL model.

In the RL model, the error structure is assumed to be IID type I extreme value, leading to MNL probabilities at each stage. For RL, Fosgerau et al. (2013) show that these value functions can be computed by solving a system of linear equations. However, little detail is presented on the complexity of this task. Fosgerau et al. (2013) note that "*for medium size networks the system [of linear equations] can be solved using a direct solution method*". Their network is small, with 7,459 links, and they can solve the problem directly, but acknowledge that for larger networks, iterative methods will be needed.

[19] Fosgerau et al. (2013) assume a large negative parameter in the utility function for any links that would imply a u-turn, thus reducing (or almost removing) the possibility of loops/cycles caused by u-turns, although of course other loops remain possible. Mai et al. (2015) estimate this penalty term rather than imposing its value, which seems preferable.

The implications of this in terms of computational requirements as well as accuracy are not immediately clear.

The resulting model structure is highly elegant and, as Fosgerau et al. (2013) demonstrate, is equivalent to the static version of a multinomial logit model, but with an infinite choice set (given the possibility for loops) of possible paths from the origin to the destination, The question in the context of the present note is how easily possible it is to operationalise the model for a large scale context such as for a national value of travel time study. The entire road network for Denmark would contain a number of links that is far larger than any previous applications of the RL model and it is not clear what the implications are in terms of computational costs.

The study by de Meyer de Freitas et al. (2019) is quite informative in this context. The authors present possibly the largest application of a RL model to date, looking specifically at multi-modal journeys, thus creating a multi-modal network of links. The authors experienced extensive issues with estimation, especially in terms of memory demands. Even after substantially reducing the complexity of the network, i.e. reducing links, the resulting model still had to be estimated on very high performance machines, reaching memory requirements of up to 80GB. Oyama & Hato (2017) similarly had to reduce the complexity of their network, and in work in Denmark[20], issues were also encountered when attempting to operationalise the model on a large scale. Of course, reducing the complexity of the network would address this issue, but this would arguably reduce the whole point of applying the model.

A number of key further developments of the RL framework have been proposed in the literature. Fosgerau et al. (2013) discuss extending the RL model to correct for overlap between routes, much as path size logit (PSL) does for non-recursive models. Standard approaches such as PSL are however not link additive. Fosgerau et al. (2013) therefore propose a different approach based on the flows on links, leading to a correction referred to as the link size (LS). The value for this LS attribute is computed using an RL model with a utility specification chosen by the analyst, rather than estimated. As the authors acknowledge, this "*utility specification is application specific and different specifications should be tested to investigate the sensitivity of the final estimation results with respect to the definition of the LS attribute*". The analyst thus needs to make assumption about the RL model parameters for simulating the flow that is then used for the LS attribute, and this is likely controversial as the final model results depend on these inputs. The results show that LS clearly improves fit, but work needs to look at the impact on relative valuations. In the Fosgerau et al. (2013) work, we see that the relative sensitivity for left turns compared to travel time drops by 16% and the relative sensitivity to crossings drops by 39% when incorporating the LS attribute. These are big changes, and work needs to establish how much they depend on assumptions made in the utility specification for the RL model used to simulate the flows. However, even with the LS, the recursive approach may remain broadly consistent with RUM.

The Mai et al. (2015) work takes the RL structure further by using nesting, i.e. moving away from the IIA assumption. The introduction of the nesting means that the value functions are now the solution of a system of non-linear equations, further complicating the structure compared to the original RL model. The authors propose an iterative approach to solve this. In illustrative examples, Mai et al. (2015) use common scale parameters for links that are in a given part of the network, and thus likely more correlated, e.g. if they have the same origin node. In practice, the number of possible ways of defining such a structure is of course unbounded and a different approach is thus needed. In any realistic network, it is impossible to estimate scale parameters for each link, and Mai et al. (2015) therefore parameterise the scale parameters as a function of the attributes of the link, with additional parameters to estimate. While this improves fit, it is not immediately clear that this captures "behavioural" correlation between routes and some further consideration is also needed in relation to the inclusion of the same attributes, especially link size, in the definition of the scale parameters and the utilities.

Another interesting contribution in this area is given by Oyama and Hato (2017). They include a discount factor so that the instantaneous utility, i.e. the "cost" of the next link, is potentially given a

[20] Private communication with Carlo Prato.

greater weight than the expected utilities of the links that follow. This is arguably more in line with sequential decision making, like in the step selection function. Using this approach, the probability of choosing a next step that involves a high cost is reduced even if the expected utility of the remainder of that path is high.

It is difficult to take a position on the suitability of RL and its extensions in the context of the present study. The model is inherently very appealing but there are questions as to its applicability in very large settings. If a reduction of the network size is needed, then we are arguably losing key benefits of a model that was put forward precisely to avoid such simplifications. Other important factors to consider relate to how easily it is to incorporate random heterogeneity in a RL model. This likely leads to further complications in solving the Bellman equations. Additionally, much of the work in route choice modelling and VTT has considered the role of non-linear sensitivities, for example in terms of cost damping. These non-linearities are more likely to apply at the overall route level rather than individual links and the non-linearity is therefore not likely to be link additive. Link additivity is an essential feature of the utility functions for applying the RL approach.

3.6. Dealing with missing information

A major advantage of SC data is that the analyst not only has extensive information about the decision maker (in terms of socio-demographics) but also about which alternatives were actually shown to be available during the choice task. In GPS data, this is not the case. The question as to the existence of consideration sets of alternatives is a controversial topic and we do not take a position on whether future work should incorporate such a component (as discussed in 3.1 above). In GPS data, especially of type 1 but also type 2, there is however additionally the problem of missing information on the actual availability of different alternatives. This relates primarily to the availability of modes of travel, i.e. is mainly a consideration issue for models of mode choice. If we disregard the possibility of mode availability varying across trips for the same traveller, which would add a further complication, then we can define π_{jn} to be the probability of mode j being available for traveller n. The inclusion of modes in a choice set is then treated probabilistically, again in a Manski-style latent class model, but in order to accommodate latent availability rather than latent consideration. When additional information is available for a traveller, this can be used to construct models of latent availability. In the example of the tagmyday study (Calastri et al., 2018b), availability of modes could be imputed on the basis of other socio-demographic information.

Even if the availability of alternatives is *known* to the analyst, many other characteristics of the traveller will be unknown, especially with type 1 and also type 2 GPS data. This relates for example to key socio-demographic measures such as income, employment and age, which are known to have an impact on VTT. If these attributes were known to the analyst, the utility conditional on these attributes could be used. Recent work by Bwambale et al (2018c) has demonstrated how models on data with missing socio-demographic information can be enriched by inferred information about such characteristics. In particular, GPS data contains information about multiple trips for each person, and the pattern of trips observed (in terms of frequency, timing, length etc) can be used to infer the likely characteristics of that person. To achieve this aim, an analyst needs to make use of separate data, for example from a source such as TU, that allows the specification of a demographic prediction model, regressing the observed traveller characteristics on the observed trip patterns. The parameters from that model can then be used to infer the probability of a traveller in the GPS data having given socio-demographic characteristics, as a function of the observed trip patterns. This thus leads to an enriched version of the GPS data, by adding imputed information about socio-demographics. The reliability of this process clearly depends on the suitability of the auxiliary data, but in the case of Denmark, the TU data would appear very suitable. Of course, these socio-demographic characteristics cannot be calculated precisely and will only be estimated up to a probability, which can then be used in the class membership function for a latent class model.

While the work by Bwambale et al (2018) used the approach to infer information about the travellers, the same process can be used for trip characteristics. A separate model would again be calibrated on data such as TU to understand the relationship between trip characteristics that are not available in the type 1 GPS data (e.g. occupancy, purpose and vehicle type, where that is not reliably available from the GPS source) and those that are (e.g. timing, length, possibly also using land use information relating to the trip's origin and destination) and the resulting model can then be used to predict this information for the trips in the GPS data.

While these methods would be essential in order to use data that was missing crucial variables, it must be said that they introduce new uncertainties into the VTT estimates.

3.7. Dealing with errors in the time and cost data

We have already referred to the need to support RP data with network-based time and cost estimates. This is often done for large-scale travel demand modelling, but the requirement there is a little less severe, as those models are applied for forecasting, using again network-based time and cost estimates. In the VTT application, we need the coefficients of the model to have a more rigorous interpretation as the true marginal disutility of travel time and cost to the traveller. The issue of bias and error in the network models therefore needs to be addressed.

Following earlier work by Walker and colleagues and by Varotto and colleagues, a recent paper by Varela et al. (2018, see that paper for details of the other references) has addressed this issue. The findings, going beyond the previous research, are that errors in both network and self-reported times and costs can be described by regression models, newly formulated to represent error as proportional to the time and cost. These regression models allow time and cost to be represented as latent variables and included in improved mode choice models. The results of the paper show that errors in cost variables are larger than errors in time variables (in the Stockholm data that was used) and hence that naïve models of this data would overstate VTT, because of greater attenuation of the cost coefficient. Elasticities would be correspondingly understated.

Clearly, account needs to be taken of this work and corresponding corrections applied to VTT estimation that is based on network estimates of travel times and costs. The Stockholm results may not be directly comparable, because for example specific network modelling software was applied, but the method that was used is applicable quite widely.

3.8. Expansion to national representativeness

In basing VTT estimates on sample data we are faced with the issues of representativeness. Issues can arise because of contact bias, response bias, or simply the random vagaries of the survey procedure. Moreover, the survey may be conducted in an area different from the area for which VTT is required (e.g. in part of the target area). A means is then required to 'expand' the survey VTT to the target area, for example to represent the whole of Denmark.

A simple method for doing this is to treat the analysis of the survey as providing a 'model' of VTT, which yields the VTT for a trip, given the purpose, mode and length of the trip, income of the traveller etc.. This model can then be applied to a sample of trips that, with their expansion factors, is representative of the target area, i.e. in a sample enumeration. This approach was applied in the UK VTT study (Arup et al., 2015) and in the recent Harbour Tunnel study. For the Harbour Tunnel study, the VTT model was applied to a sample of TU trips representative of the Copenhagen area but of course it could also be applied to a TU sample representative of Denmark as a whole.

The sample enumeration approach means that biases in the sample are corrected with respect to variables that are included in the model and in the expansion data. We cannot correct for biases with respect to variables that are not included in both data sources. For example, the attitude to congestion may be different between Copenhagen and Jylland, but this cannot be measured and therefore the application of the Harbour Tunnel SC models for the whole of Denmark is not advisable.

A further advantage of the sample enumeration process is that error measures can be calculated. These relate to two potential sources of sampling error:

- error in the estimation of the parameters, e.g. as indicated by the standard errors given by the estimation procedure, which can be applied using the 'delta' method (Daly et al., 2012);
- sampling error in the sample enumeration sample, which can be estimated by a bootstrap method.

These procedures were applied in the UK VTT study (Arup et al., 2015) but not in the Harbour Tunnel study, where time and resource constraints prevented it.

At the same time, adjustments can be made to the year for which VTT is required. Note that the enumeration sample may itself have a different year from the VTT sample and the required output year may be different again. For very short periods (e.g. 1-2 years) simple adjustments for inflation can be made, but over longer periods it may be necessary to take account of changes in 'real' income levels. These adjustments can be quite complicated.

4. Summary and conclusions

In Denmark, as in several other countries, SC methods have been used to estimate travellers' WTP for travel time. However, a series of problems remain with the SC approach and it is reasonable to consider what alternative methods could be used. In particular, the increasing availability of 'large data' suggests that GPS or other RP approaches could offer an improvement over SC. RP was the original approach to VTT estimation but was largely abandoned from the 1990s onward.

At the same time, it is worth considering whether the WTP approach should also be used for estimating VTT of travellers who are working. CSA has been used in most countries, including Denmark, for working-time VTT but it has its own limitations and if improved WTP approaches are available this might change the balance. The UK has recently made this change.

4.1. Summary on RP data

A key issue in using RP data is to decide which of the traveller's choices should be modelled. Route choice has typically been used, but in RP data the variances of time and cost are low across routes for the same mode. Mode choice then becomes an obvious candidate, although issues of availability of modes and non-measurable preferences complicate the modelling. Destination choice or departure time choice present other options, also with associated difficulties.

Conventional RP data takes the form of trip diaries, either for a single trip, for a day (as in the case of the Danish TU data) or for longer. Such data can be used for VTT estimation and is available with little additional cost.

In the case of GPS data, three broad forms can be identified:

1. simple loggers installed in vehicles (as used in the data analysed by DTU for the Harbour Tunnel study);
2. GPS loggers carried by survey respondents, who are asked to add (often only some basic) information;
3. data collected on a smartphone carried by respondents, who are asked to add information about trips; these surveys also collect all the background data typically collected in household travel surveys.

An overview of the key properties of these data types and a comparison with SC and TU data is given in the table at the end of this note.

In the Harbour Tunnel study, the difficulties encountered by DTU are characteristic of GPS type 1 data, in that the cost coefficient was difficult to estimate and eventually a generalised cost approach had to be used. In other studies, monetary VTT values have been estimated, but with difficulty. It seems that data of this type cannot reliably be used on its own for monetary VTT estimation, especially when focusing on route choice or departure time choice. More scope may exist for estimating non-monetary VTT aspects, such as congested vs. non-congested time, or schedule delay vs. travel time.

Some success has been obtained with GPS data of type 2, but this is now being replaced by type 3 data. Type 3 data has the advantage over type 2 that respondents are less likely to forget their phone than a separate logger. Moreover, the app that is used can remember locations, meaning that less information has to be entered by the respondent over time, improving response.

Another big data source that can be considered is the use of mobile phone records. This requires negotiation with the relevant companies and appropriate guarantees of privacy, but can yield huge quantities of data. Of course, it is less useful for modelling that the type 2 or 3 GPS data mentioned above.

With GPS and phone data, data processing becomes an issue. It is necessary to identify origins and destinations of trips and to rectify inaccuracies of the location coding provided automatically.

For all kinds of RP data, it is necessary to supply travel time and cost estimates. Reported values should not be used and this implies either the use of a network model or using a less systematic method of looking up the relevant values for a trip. Correction procedures for networks have been shown to be effective in reducing error.

In particular, trip costs pose a problem. When marginal costs are used in appraisal, we would recommend the use of marginal costs in VTT estimation, to maintain consistency with the economic framework of appraisal. However, in Denmark, average costs are used in the appraisal process, which could suggest the use of average costs in VTT estimation, though this presents issues for modelling. For travel in the course of work, the marginal cost to the business may be the overall average cost. For walking and cycling, it has proved difficult to introduce trip costs, so that mode choice against modes that do have costs is recommended as a means of estimating VTT for these modes.

Joint use of data of several types may be useful, with GPS data of type 3 offering the most effective current approach of obtaining rich information about trips and travellers.

If VTT is required for AV, it seems there is no alternative to SC methods. However, these would depend on a careful explanation to potential travellers of the circumstances of their hypothetical journey and detailed consideration of potential biases. A reasonable approach might also be to imagine that the VTT with AV is similar to the VTT with taxis now.

4.2. Summary on estimation

The overall framework in which estimation is done should respect the RUM paradigm, since this is consistent with the economic framework in which appraisal is to be performed. RUM can include a number of flexible representations of behaviour, although path size logit (often used in GPS route choice modelling) is not totally consistent with RUM (for reasons explained in detail in 3.4). Detailed studies need to be made of the various forms of heterogeneity present in the population.

With GPS and other RP data, the focus on route choice for VTT estimation needs to be reconsidered. For type 1 GPS data, other choices are difficult to model, so consideration of type 3 data becomes more attractive; the availability of trip purpose and income data is another important issue here. Other RP data, such as TU trip diaries, can also be considered, where modelling mode and destination choice becomes a possibility. Network models and their correction are required for using RP data.

While the use of RP data means that the chosen alternative is well defined, the issue of specifying which alternatives were rejected is quite serious. The number of alternatives that might feasibly have been considered can be very large in the case of destination choice or route choice, although the route choice situation can be alleviated by adopting a recursive approach as discussed below. Reducing the choice set by considering behavioural issues (i.e. finding the consideration set) or by randomly sampling the alternatives might either reduce or increase bias. Either way, substantial complexities may be introduced into the modelling process, mainly due to the corrections required to reduce/avoid bias. Testing is also necessary to determine to what extent the VTT estimates are affected by assumptions made in the specification of the choice set.

It seems reasonable to assume that choices are made at the outset, i.e. as if travellers have knowledge of the conditions they will encounter along their journeys. A key issue is then how commonality or overlap between alternatives can be represented, which is particularly relevant when route choice is being modelled, involving physical overlap of the alternatives on shared links. Two methods that have been used to represent overlap, C-Logit and Path Size Logit, appears to work well in representing behaviour but are not fully consistent with RUM and therefore raise issues when considering their use for VTT estimation.

A possible approach for route choice modelling and which could avoid sampling of alternatives in this context is 'recursive logit'. However, the application of recursive logit at national scale for Denmark raises substantial computational issues, which do not seem to have been entirely solved by recent leading-edge research. We conclude that modelling route choice involves either testing of choice set selection procedures or involvement in new research.

The use of some forms of GPS data raises the issue of missing data relating to key variables, such as travel purpose, traveller's income etc.. Methods exist for inferring the missing data, using relationships in the TU data and the variables that are observed for each trip, such as length and timing. However, these methods would obviously give VTT estimates that are less reliable than if the data could be observed directly.

In any VTT estimation based on network models, correction procedures for the time and cost variables generated by those models need to be applied.

A sample enumeration procedure using TU data should be used to expand the VTT estimated from sample data to national totals and to estimate errors in the national estimates. This procedure will eliminate some but not all biases in the estimates.

4.3. Evaluation and validation

The preceding sections have discussed *how* RP data could be used to estimate VTT; here we look at the issue of *whether* RP is a good approach, given the existing and widely-used alternative of SC data. It is clear from our discussion that RP data has limitations, in particular that it has not been widely used for VTT estimation and so is not thoroughly validated, but we have noted that SC data raises a number of important and unresolved issues also. A decision to use RP would be based on a balanced judgement.

We have discussed a number of forms of RP data, which each have strengths and weaknesses. We identified three forms of GPS data, to which we add mobile phone records and traditional survey data like the TU data. A comparison of these data types with SC data is given in the table at the end of the note.

- The simple GPS data derived from vehicle-based loggers, which we called type 1, as used by DTU in their work for the Harbour Tunnel study, does not seem to offer much advantage for VTT estimation:
    - it can be used only for route choice or (possibly) time period choice, which offer very limited cost variation but would allow modelling of the additional disutility of congestion;
    - it does not contain socio-economic data or key trip data (such as purpose and occupancy) so that these have to be inferred, with error, from other information;
    - extensive and error-prone data processing is needed;
    - on the other hand, large quantities of data, possibly over repeated days, can be obtained at little marginal cost;
    - it offers excellent possibilities for obtaining relative values, e.g. for congestion and free-flow time; and

    - substantial quantities of this data exist already.

- GPS data derived from a mobile phone app (which we called type 3, superseding type 2) appears to offer more advantages:
    - locational and timing data is comparable with type 1 or possibly better, given that respondents are given the option of "correcting" the recorded information;
    - users can be asked to enter specific information about themselves and their trips, using the intelligence of the app to keep the requirement for user input to a minimum; however, some failures and refusals will occur and these will reduce the value of the data;
    - this approach is not cheap, as the app has to be prepared (in Danish) and respondents have to be recruited;
    - information concerning many trips can be obtained from each respondent, potentially covering several modes and destinations.

- Mobile phone records suffer from all of the deficiencies of type 1 GPS data. They also require negotiation with the phone companies and may give less accurate locations than GPS loggers. This data seems more suited to deriving a picture of overall traffic flows than trip specific decisions such as are needed for VTT estimation.

- Traditional survey data like the TU data can also be used usefully to model mode choice or mode and destination choice.
    - Like all RP models using network time and cost measures, correction procedures need to be applied but then we have confidence that the time and cost parameters can be estimated accurately.
    - In any case, the TU data should be used to expand the VTT model to give nationally representative VTT.

We conclude that the best information is likely to be obtained from the type 3 data using an app. Given the low additional cost of type 1 and TU data, we recommend the exploration of joint estimation using these three sources. This could be conducted by maximising likelihood over several data sources simultaneously.

We do not recommend joint estimation using RP and SC data, as this would imply a national SC data collection exercise, which would be expensive. Instead, the existing SC data from the Harbour Tunnel project should be compared with the new RP values, restricting comparison to the Copenhagen area, but remembering always that the SC values themselves are subject to concerns. For estimating VTT for AV, a new SC exercise seems inevitable, however, the alternative being an assumption that the values are comparable to those for (e.g.) taxi.

For the main VTT estimation, we recommend the modelling of mode choice (or possibly mode and destination choice), rather than route choice or departure time choice, as route and departure time choice do not give sufficient variation in cost for reliable VTT estimation. Existing GPS data could be used to model route choice (and possibly departure time choice) as the costs of this data are low and relative values can be effectively estimated.

Travel time and marginal cost estimates for use with RP data should be derived from network models, with a correction procedure to account for errors in those variables.

It is essential to formulate models within the RUM paradigm and this excludes some specifications of C-Logit and Path Size logit (if time and cost enter the "length"), further reducing the case for using route choice for VTT estimation. As discussed, recursive choice models do not seem a complete solution to the problems either, at least at the current stage of their development. However, it seems reasonable to formulate models of behaviour 'as if' travellers had full information about possible choices at the outset. Defining the choice set is a key step and this is much more easily done in mode choice than in other traveller choices. Availability and consideration sets can also be modelled when the full set of feasible alternatives is moderate.

VTT models can be expanded to national representativeness using the sample enumeration procedure. This will also allow error measures to be created. It is interesting to compare these with those derived from SC models. Our anticipation would be that the RP models would be at least as precise, according to these measures, as the SC results. Of course, precision depends on the number of observations available for modelling, so that a good idea could be obtained by comparing results of OTM or LTM modelling with the Harbour Tunnel SC models. In either case, a square root law would apply to errors, so the sample size would have to be quadrupled in order to halve the error. Theoretically, data should be spread over the year, but practicalities of project schedules and budgets may make this impossible.

Precision, however, is not the key issue; rather we are concerned with bias. In principle, RP data is less subject to bias than SC data. Validation of VTT results is difficult, because of the lack of precisely comparable sources, but meta-analysis work (e.g. Wardman et al., 2016) can give some reassurance that results are not too far out of line with other European findings; these of course are also subject to bias, though Wardman includes both RP and SC values in his analysis. Wardman's work also points to the variables that have been found to be most important in determining VTT.

The great advantage of RP compared with SC data is that it does not suffer from hypothetical bias and this means that a careful investigation of the RP possibilities is worthwhile.

References

Antonisse, R.W., Daly, A. and Ben-Akiva, M. (1988) A Highway Assignment Model based on Behavioural Models of Drivers' Route Choice, *Transportation Research Record 1220*.

Arup, ITS and Accent (2015) Provision of market research for value of travel time savings and reliability, Phase 2 Report, for DfT
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/470231/vtts-phase-2-report-issue-august-2015.pdf
Accessed 5 Nov 2018

Axhausen, K.W., Zimmermann, A., Schönfelder, S., Rindsfüser, G. & Haupt, T. (2002), Observing the rhythms of daily life: A six-week travel diary, Transportation, 29(2), pp.95-124.

Axhausen, K. W., I. Ehreke, A. Glemser, S. Hess, C. Joedden, K. Nagel, A. Sauer and C. Weis (2015) Ermittlung von Bewertungsansätzen für Reisezeiten und Zuverlässigkeit auf Basis der Schätzung eines Modells für modale Verlagerungen im nicht-gewerblichen und gewerblichen Personenverkehr für die Bundesverkehrswegeplanung: FE-Projekt 96.996/2011 Zeitkosten Personenverkehr, Research Report, Federal Ministry of Transport and Digital Infrastructure (BMVI), IVT, ETH Zurich, Berlin.

Bellman, R., 1957. Dynamic Programming. Princeton University Press, Princeton.

Ben-Akiva, M.E., Bierlaire, M. (1999), Discrete choice methods and their applications to short term travel decisions. In: Hall, R.W. (Ed.), Handbook of Transportation Science. Kluwer, Dordrecht, The Netherlands, 5-34.

Bwambale, A., Choudhury, C.F. and Hess, S., (2018a), Modeling Long-Distance Route Choice Using Mobile Phone Call Detail Record Data: A Case Study of Senegal. The Transportation Research Board (TRB) 97th Annual Meeting, Washington, D.C.

Bwambale, A., Hess, S., Choudhury, C. (2018b), Analysing time-of-travel choices using mobile phone data, 15th International Conference on Travel Behaviour Research, Santa Barbara

Bwambale, A., Choudhury, C.F. & Hess, S. (2018c), Modelling trip generation using mobile phone data: a latent demographics approach, Journal of Transport Geography, forthcoming.

Calastri, C., Crastes dit Sourd, R. (2018a), We want it all: experiences from a survey seeking to capture social network structures, lifetime events and short-term travel and activity planning, Transportation, forthcoming.

Calastri, C., Hess, S., Choudhury, C.F., Daly, A.J. & Gabrielli, L. (2018b), Mode choice with latent availability and consideration: theory and a case study, Transportation Research Part B, forthcoming

Cascetta, E., Nuzzolo, A., Russo, F., Vitetta, A. (1996), A modified logit route choice model overcoming path overlapping problems: specification and some calibration results for interurban networks. In: Lesort, J.B. (Ed.), Proceedings of the Thirteenth International Symposium on Transportation and Traffic Theory. Pergamon, Lyon, France, 697-711.

Cascetta, E. & Papola, A. (2001), Random Utility Models with Implicit Availability/Perception of Choice Alternatives for the Simulation of Travel Demand, Transportation Research Part C 9 (4), pp. 249-263.

Daly, A., Hess, S. and de Jong, G. (2012) Calculating errors for measures derived from choice modelling estimates, Transportation Research B, Vol. 46, pp. 333-341.

DfT (2017) WebTAG advice for transport planners,
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/603254/webtag-tag-unit-a1-3-user-and-provider-impacts-march-2017.pdf
Accessed 1 Nov 2018

Dial, R. (1971) A probabilistic multipath traffic assignment model which obviates path enumeration, Transportation Research, 5, pp. 83-111.

Dijkstra, E.W., 1959. A note on two problems in connexion with graphs. Numer. Math. 1, 269–271.

Fosgerau, M., Bierlaire, M., (2009), Discrete choice models with multiplicative error terms.,Transp. Res. Part B, 43, pp. 494–505.

Fosgerau, M., Frejinger, E. & Karlstrom, A. (2013), A link based network route choice model with unrestricted choice set, Transportation Research Part B, 56, pp. 70-80.

Fosgerau, M. , Mabit, S. , (2013), Easy and flexible mixture distributions, Econ. Lett, 120, pp. 206–210.

Guevara, C.A. and Ben-Akiva, M. (2013), Sampling of alternatives in multivariate extreme value (MEV) models, Transportation Research Part B 48, pp. 31–52

Harrison, G. (2014), Real choices and hypothetical choices, in Hess, S. & Daly, A.J. (eds), Handbook of Choice Modelling, Edward Elgar Publishing, Cheltenham, UK.

Hess, S., Quddus, M., Rieser, N. & Daly, A.J. (2015), Developing advanced route choice models for heavy goods vehicles using GPS data, Transportation Research Part E, 77, pp. 29-44.

Hess, S., Daly, A., Dekker, T., Ojeda Cabral, M. and Batley, R. (2017) A framework for capturing heterogeneity, heteroskedasticity, non-linearity, reference dependence and design artefacts in value of time research, Transportation Research B, Vol. 96, pp. 126-149, Feb.

Hess, S., Daly, A. and Batley, R. (2018) Revisiting consistency with random utility maximisation: theory and implications for practical work, Theory and Decision, January.

Frejinger, E. & Bierlaire, M. (2007), Capturing correlation with subnetworks in route choice models, Transportation Research Part B 41 (2007) 363–378.

Frejinger, E., Bierlaire, M., and Ben-Akiva, M. Sampling of alternatives for route choice modeling. Transportation Research Part B, 43(10), pp. 984–994, 2009.

Fortin D, Beyer HL, Boyce MS, Smith DW, Duchesne T, Mao JS (2005), Wolves influence elk movements: behavior shapes a trophic cascade in Yellowstone National Park. Ecology, 86, pp. 1320–1330.

Mai, T., Fosgerau, M. & Frejinger, E. (2015), A nested recursive logit model for route choice analysis, Transportation Research Part B 75 (2015) 100–112

Manski, C. F. (1977) The structure of random utility models. Theory and decision, 8(3), 229-254.

McFadden, D. & Train, K.E. (2017), Contingent Valuation of Environmental Goods: A Comprehensive Critique, Edward Elgar Publishing, Cheltenham, UK.

Meyer de Freitas, L., Becker, H., Zimmermann, M. & Axhausen, K.W. (2019), Modelling intermodal travel in Switzerland: A recursive logit approach, Transportation Research Part A: Policy and Practice, 119, pp. 200-213.

Montini, L. Prost, S., Schrammel, J. Rieser-Schüssler, N. & Axhausen, K.W. (2015), Comparison of Travel Diaries Generated from Smartphone Data and Dedicated GPS Devices, Transportation Research Procedia, Volume 11, pp 227-241.

Oyama, Y. & Hato, E. (2017), A discounted recursive logit model for dynamic gridlock network Analysis, Transportation Research Part C 85 (2017) 509–527.

Paag, H., Daly, A.J., Rohr, C. (2001) Predicting use of the Copenhagen harbour tunnel, pp. 627-646 in *Travel Behaviour Research: the Leading Edge*, David Hensher (ed.), Pergamon.

Prato, C. G. (2009), Route choice modeling: past, present and future research directions, Journal of Choice Modelling, Volume 2, Issue 1, pp. 65-100.

Swait, J. (2001) Choice set generation within the generalized extreme value family of discrete choice models. Transportation Research Part B, 35(7), pp. 643-666

Train, K. & Weeks, M. (2005), Discrete Choice Models in Preference Space and Willingness-to-Pay Space, Ch. 1, pp. 1-16, in Applications of Simulation Methods in Environmental and Resource Economics, A. Alberini and R. Scarpa, (eds)., Springer Publisher: Dordrecht, The Netherlands, 2005.

Varela, J.M.L., Daly, A. and Börjesson, M. (2018) Quantifying errors in travel time and cost by latent variables, Transportation Research B, 117, pp. 520-541.

Wardman, M. et al. (2013) VALUATION OF TRAVEL TIME SAVINGS FOR BUSINESS TRAVELLERS, ITS Report for DfT, April 2013
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/251997/vtts_for_business_main_report-dft-005.pdf .
Accessed 1 Nov 2018

Wardman, M., Batley, R., Laird, J., Mackie, P. and Bates, J. (2015) How should business travel time savings be valued?, Economics of Transportation, 4, 200–214.

Wardman, M, Chintakayala, V.P.K. and De Jong, G.C. (2016) *Values of travel time in Europe: Review and meta-analysis.* Transportation Research Part A: Policy and Practice, 94. pp. 93-111. ISSN 0965-8564

Overview of main characteristics of data sources

<table>
<tr><th></th><th>SC</th><th>Travel diary data (e.g. TU)</th><th>Type 1 GPS data</th><th>Type 2 GPS data</th><th>Type 3 GPS data</th></tr>
<tr><td>Journey and traveller characteristics</td><td>Collected to level of detail specified by analyst</td><td>Collected to level of detail specified, usually by others
Most freight excluded</td><td>Not generally available, some, like vehicle type, may be imputed</td><td>Available to a limited extent in some surveys, but missing data may be an issue</td><td>Collected to level of detail specified by analyst, data of quality at least as high as with traditional diary data</td></tr>
<tr><td>Representativeness</td><td colspan="2">Controlled by analyst, in terms of socio-demographic representativeness, but not necessarily VTT representativeness</td><td>Large sample may help, but generally no way to check</td><td colspan="2">Controlled by analyst, in terms of socio-demographic representativeness, but not necessarily VTT representativeness</td></tr>
<tr><td rowspan="2">Repeated observations</td><td rowspan="2">Yes, typically limited to 10-16 per person. With variations across observations, but generally relating to the same base reference trips</td><td rowspan="2">Very limited, although some surveys collect data on multiple days</td><td colspan="3">Generally many observations per respondent, often covering multiple different trip lengths per person</td></tr>
<tr><td>Limited or no variations across similar journeys, given unimodal context

Links across observations for the same respondent may be hidden for privacy reasons (IDs scrambled)</td><td colspan="2">Variation across days and across modes, though within mode/route variation may be limited at the respondent level</td></tr>
<tr><td>Trade-offs to assist VTT calculation</td><td>By design</td><td colspan="4">Limited by the trade-offs existing in real life, but easier if we include more choices (e.g. mode, destination)</td></tr>
<tr><td>VTT in broader activity context</td><td>Choices in isolation, no consideration of impact on other activities</td><td>Some activity information is available</td><td colspan="2">Real world choices, with impact on other activities, but those are not usually observed</td><td>Real world choices, with impact on other activities, which can be captured too</td></tr>
<tr><td>Choice set</td><td>Predetermined by analyst, but with obvious endogeneity risks</td><td>Some conditioning variables are observed (e.g. car ownership)</td><td>Unobserved</td><td colspan="2">Unobserved, but with potential background questions (also with possible endogeneity)</td></tr>
</table>