HOW LIKELY ARE TRAVELERS TO GIVE UP INFORMATION IN EXCHANGE FOR BETTER USER INFORMATION SERVICES?

Constantinos Antoniou*
Assistant Professor, Laboratory of Transportation Engineering
School of Rural and Surveying Engineering
National Technical University of Athens
9, Iroon Polytechniou St., Zografou Campus, 15780
Tel: 0030-210-7722783; Fax:0030-210-7722629
Email: antoniou@central.ntua.gr

Amalia Polydoropoulou
Professor
Department of Shipping Trade and Transport
University of the Aegean
Korai 1A, 82100, Chios, Hellas
Email : polydor@aegean.gr

*Corresponding author

Word Count: 6250 words + 3 tables + 2 figures = 7500

Submitted on August 1st, 2013; Revised version submitted on November 6th, 2013
Submitted for presentation in the 93rd Annual Meeting of the Transportation Research Board.

TRB 2014 Annual Meeting Paper revised from original submittal.
HOW LIKELY ARE TRAVELERS TO GIVE UP INFORMATION
IN EXCHANGE FOR BETTER USER INFORMATION SERVICES?

ABSTRACT
The emergence of location-based services has raised privacy to a hot-issue status. Clearly, there is a trade-off that needs to be considered. On the one hand, additional information can arguably improve the quality of the various services, not only for the individual, but also for the entire system. On the other hand, the collection of this additional information can be considered by some as a violation of their privacy.

The objective of this paper is twofold: on the one hand to provide a way to quantify the value of this privacy (loss) and on the other hand to present results from a case study applying this methodology. The methodology is generic and is based on the concept of the marginal rate of substitution between cost and a different variable (in this case privacy). The application is based on a stated-preference questionnaire that was specifically designed for this application and disseminated to respondents in Athens, Greece.

The estimated willingness-to-accept for giving up one level of privacy is equal to 2.2€/month (approx. US$2.8/month), which is considered reasonable since such services are commercially available for US$5/month. Results from a gender-based market segmentation analysis indicate that women appear to be more sensitive to privacy, but also show a slightly higher willingness-to-pay for improved accuracy.

Keywords: privacy, stated-preference survey, random-effects ordered probit model, marginal rates of substitution, willingness to pay
INTRODUCTION

Vehicles are currently equipped with a large number of on-board computers, as well as communication capabilities enabling the exchange of information between vehicles (V2V) and between the vehicle and infrastructure (V2I), thus paving the way for a connected system that can result in very rapid collection of information (1). A large number of sensors are also available on or around the road, complementing this mesh of data and completing the coverage. Antoniou et al. (2) present a synthesis of emerging data collection technologies, including GPS and opportunistic sensors, as well as their impact on traffic management applications.

The use of smartphones and related applications for transportation and traffic management applications is also the topic of several large-scale and high-profile research projects, such as the mobile millennium project, a UC Berkeley, Caltrans, Nokia, NAVTEQ collaboration, that includes a pilot traffic monitoring system that uses the GPS in cellular phones to gather traffic information, process it and disseminate it back to the nomadic devices (http://traffic.berkeley.edu, (3)). The SMART (Singapore-MIT Alliance for Research and Technology) initiative also includes several projects leveraging information from nomadic devices for collection of information and the reduction of energy use (http://smart.mit.edu/).

Technologies to process large volumes of location sensor data are being developed, allowing real-time traffic information management (4).

Clearly, there is a trade-off that needs to be considered. On the one hand, additional information can arguably improve the quality of the various services, not only for the individual, but also for the entire system. On the other hand, the collection of this additional information can be considered by some as a violation of their privacy. In order to be able to assess whether it is worth trying to obtain this information, an entity -that e.g. needs additional personalized information to develop a new service- would like to know the cost associated with obtaining this information. This cost may be directly monetary, or perhaps associated with the perception of this service or the willingness of the people to use it. In any case, in order to be able to assess it, it needs to be converted to a single unit, and this is usually monetary units. Provided such estimate, a value-added provider or systems developer can choose whether the cost of obtaining the additional information (in order to provide a superior service) is justifiable and makes business sense.

The objective of this paper is twofold: on the one hand to provide a way to quantify the value of this privacy (loss) and on the other hand to present results from a case study applying this methodology. The methodology is generic and is based on the concept of the marginal rate of substitution between cost and a different variable (in this case privacy). The application is based on a stated-preference questionnaire that was specifically designed for this application and disseminated to respondents in Athens, Greece.

The remainder of this paper is structured as follows. The next Section provides relevant background and places this research in the context of the literature and the state-of-the-art. The following Section presents the methodological background, in terms of the econometrics models and methods that are used to quantify the perceived loss of privacy. The case study setup is described in the next Section, in terms of the questionnaire design and data collection effort, as well as the general characteristics of the responses. The results of the estimated models are presented next, followed by a gender-based market
segmentation analysis. The next section provides an application of the model estimation results, while the concluding Section provides additional insight and outlines directions for future research.

BACKGROUND

Policy aspects
Cottrill and Thakuria (5) conclude that “more consistent regulatory guidance would be useful to ensure that there is adequate protection of consumer location data” and that the public would feel better protected if a guarantee of legal protection were provided. Leduc (6) presents an interesting discussion about the main concerns related to privacy issues, including the type of data that is being collected, by whom and for how long they are retained. Leduc (2008) also discusses the issue of data ownership and sharing, which are time-sensitive as already there is a critical mass of such data being collected.

Personal data protection requires policy intervention, while the international nature of the issue makes the need for international standards pressing. Indeed, international standards are being developed (ISO/TC204/WG16), but considering the discrepancy between the glacially slow process of developing standards and the speed at which developments in the field of tracking occur, it is evident that more effort is needed. Currently, the “benchmark” regulation for data protection is the European Union Data Protection Directive (Directive 95/46/EC), which defines personal data rather broadly and defines the conditions under which the data can be collected, processed and shared.

Quantifying the value of privacy
In a very interesting study, Cottrill and Thakuria (7) used data from an online survey questionnaire to investigate perceived privacy risks. The expected compensation requirements for the “breach” of privacy did not change in most of the considered scenarios. However, in the case of the information leading to a decrease in road accident fatalities by 1000 persons per year, a higher percentage of the respondents indicated that their compensation would decrease. On the other hand, when considering information collected by agencies in order to be sold to third parties, the majority of respondents reported that their compensation requirements would increase.

In a different context, Hann et al. (8) use questionnaire data from the US and Singapore in which they offer monetary rewards for certain breaches of online privacy. Hann et al. (8) use the marginal utilities (from conjoint analysis models they estimate) of the monetary reward and the “part-worths” for privacy protection to estimate the value of protection for each of three privacy concerns (“review for error”, “restriction against improper action”, “secondary use not allowed”).

Grossklags and Acquisti (9) propose two interesting variants to the concept of willingness-to-pay, in particular “willingness-to-protect”, i.e. willingness-to-pay for protecting information and “willingness-to-sell”, i.e. willingness-to-accept a proposal to sell information. The authors developed a two-stage experiment to both evaluate the response of the subjects to fixed offers, but also obtain information on the maximum stated values for their willingness-to-protect/accept.

Guderian (10) identifies privacy issues as one of four main barriers to a mileage-based tax system (along with equipment reliability, support of policy makers and implementation
process). Similarly, Sana et al. (11) consider privacy aspects of a vehicle-mileage based user fee. Cruickshanks and Waterson (12) present the results of a survey in the UK, from which they conclude that there are fears that some future applications may cause some people to travel with less freedom, which can have a negative impact to the society as a whole. It is noted, however, that respondents’ privacy concerns towards the use of GPS in travel surveys have been minimal (13, 14).

**METHODOLOGY**

**Ordered probit model**

Respondents in surveys are often asked to express their preferences in a rating scale. Such scales are often called Likert scales (15, 16). A multinomial logit model could be specified with each potential response coded as an alternative. However, the ordering of the alternatives violates the independence of the errors for each alternative, and therefore the Independence for Irrelevant Alternatives (IIA) assumption of the logit model. Nested or cross-nested models are one approach to overcoming this issue (17), while multinomial probit models also do not suffer from this limitation. Ordered logit and probit models provide another approach that estimates parameter coefficients for the independent variables, as well as intercepts (or threshold values) between the choices.

Assuming a ranking scale with seven levels, there are six thresholds or critical values (k1 through k6) that separate the seven choices (“Certainly A”, “Probably A”, “Possibly A”, “Indifferent”, “Possibly B”, “Probably B”, “Certainly B”). For example, respondents choose the alternative “Certainly A” if the utility is below k1, alternative “Possibly A” if the utility is between k1 and k2, and so on.

In an ordered probit model, the ordered response is used directly as the dependent variable. In each model, the response variable takes numerical values between 1 and 7, with 1 indicating that the respondent is stating that he would certainly choose alternative A and 7 indicating that the respondent would certainly choose alternative B.

If Y is the response factor with K levels, the model can be written as:

\[ P(Y \leq k|x) = \Phi (\theta_j - \beta^T x) \]

where:

- \( \Phi \) is the cumulative normal function,
- \( \theta_0 = -\infty < \theta_1 < \cdots < \theta_k = \infty \) are the breakpoints,
- \( x \) is the vector of the explanatory factors, and
- \( \beta \) is the vector of the unknown parameters.

**Random effects models**

As mentioned above, the data used in this research involve repeated observations from each individual. When dealing with such panel data it is often useful to consider the heterogeneity across individuals, often referred to as unobserved heterogeneity. In general, pooling data across individuals while ignoring heterogeneity (when it is present) will lead to biased and inconsistent estimates of the effects of pertinent variables (18). Several approaches have been developed to incorporate these effects in the model formulation.
One such approach is to estimate a constant term for each individual and each choice, which is referred to as a "fixed-effects" approach (19). Perhaps the main drawback to this approach is the large number of parameters (and consequently large number of required observations per individual). A more tractable approach is to assume that the fixed term varies across individuals according to some probability distribution, which is referred to as a “random effects” specification (18, 20).

**Indifference curves and the marginal rates of substitution**

Trade-offs such as these are not uncommon in practice and—as expected—microeconomic theory includes suitable tools to treat them: indifference curves and the marginal rate of substitution (MRS), discussed next following Nicholson (21). An indifference curve is defined as a curve \( U_1 \) that represents all the alternative combinations of two goods \( X \) and \( Y \) for which an individual is equally well-off (assuming that all other elements of the utility function are kept constant). The slope of the indifference curve is negative, showing that if the individual gives up some of good \( Y \), then in order to reach an equivalent bundle of goods (staying on the same indifference curve \( U_1 \)) the individual needs to be compensated by an additional amount of \( X \). The changing slope of the curve indicates the assumption that individuals become progressively less willing to trade more units of \( Y \) in order to obtain additional units of \( X \).

This can be formalized through the concept of marginal rate of substitution, i.e. the negative of the slope of an indifference curve (\( U_1 \)) at some particular point:

\[
MRS = -\frac{dY}{dX} \bigg|_{U=U_1}
\]  

(1)

Where the notation indicates that the slope is to be calculated along with the \( U_1 \) indifference curve.

Suppose the following general formulation for the systematic component of the utility function is used:

\[
V = \beta_0 + \beta_{\text{cost}} * \text{cost} + \beta_{\text{privacy}} * \text{privacy} + \beta_{\text{accuracy}} * \text{accuracy} + \ldots
\]  

(2)

where \( \beta_i \) are the coefficients to be estimated, \( \text{cost, privacy and accuracy} \) are the variables associated with cost of use of the service, level of operation with respect to privacy and service accuracy and “…” corresponds to additional explanatory parameters in the model. The coefficient of the cost and the coefficient of the privacy level capture the sensitivity of the travelers' utility towards changes in the accuracy and the cost. Their ratio can therefore be used to capture the trade-off between the privacy level and the cost of the service; in other words a measure of the value-of-privacy. The following explanation provides some more insight into this. The utility is in general unitless. To simplify notation, it is sometimes useful to express it in an imaginary unit of “utils”. Assuming that the cost of the service is measured in Euros (€, approx. US$1.3 in July 2013) and the privacy has a number of pre-defined levels, the units of the respective coefficients could then be utils/€ and utils/level of privacy respectively. The ratio of the coefficient for the privacy level over the coefficient for the monthly service cost would have units of €/privacy level, which is the expected unit for a value-of-privacy (VOP) measure:
Similarly, the value of service accuracy (VOA) can be obtained by:

\[
V_{OA} = \beta_{\text{accuracy}} \left( \frac{\text{utils} / \text{accuracy level}}{\text{utils} / \€} = \€ / \text{accuracy level} \right)
\]  

(4)

One assumption that is being implicitly made here is that the value of going from one privacy level to the next (respectively from one accuracy level to the next) is constant, irrespective of the starting level of privacy (or accuracy). This assumption might be particularly restrictive in the case of a revealed preference experiment, in which the input is a specific service. However, in the case of stated-preference experiments, the researchers can take care to design an experiment that satisfies this requirement.

**CASE STUDY SETUP**

**Questionnaire design and data collection**

In order to conduct the stated-preference survey, a three-part questionnaire was designed. In the first part, there were some general questions on the participant’s driving experience, use of travel modes and familiarity and use of GPS navigation devices. The second part included the actual choice experiment and the third part aimed at the collection of socioeconomic data about the respondent.

Before the participants filled-in the questionnaire, they were asked to read the instructions and learn about the general goal of the survey. In order to minimize various forms of response and justification bias, the stated objective of the survey was listed as collection of data prior to the introduction of a new satellite navigation service that has the capability to exploit additional data about the user in order to provide more accurate routing.

The statistical design for the choice experiment asked the user to indicate their preference for each of two hypothetical satellite routing services, considering three fundamental variables for the service in question:

a. Cost (€/month): for the use of the service;

b. Accuracy: (i) good, i.e. a reference level of usual navigation devices; and (ii) better, i.e. a level of improved performance, due to the utilization of additional information; and

c. Way of operation: (i) without any additional information about the traveler; (ii) with the location and the destination of the traveler; and (iii) with information about the driving pattern of the traveler.

A fractional factorial design (22) was selected from the full design. A number of 8 scenarios were then selected to be included in the questionnaire. The two alternatives in each choice were then presented to the respondents, who were asked to rate their preference for either one in a seven-level rating scale (15). Figure 1 presents a sample scenario, as presented to the survey participants.
Which alternative route would you choose?

<table>
<thead>
<tr>
<th></th>
<th>Alternative A</th>
<th>Alternative B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mode of operation</td>
<td>No location data</td>
<td>Driving pattern</td>
</tr>
<tr>
<td>Accuracy</td>
<td>Better</td>
<td>Good</td>
</tr>
<tr>
<td>Cost</td>
<td>8€/month</td>
<td>2€/month</td>
</tr>
</tbody>
</table>

Strong preference towards A  Moderate preference towards A  Mild preference towards A  Indifferent  Mild preference towards B  Moderate preference towards B  Strong preference towards B

FIGURE 1. Sample scenario

Data collection and analysis

The developed questionnaire was disseminated through face-to-face interviews. In the interviews, the respondents were handed a hard copy of the questionnaire and were asked to fill it out according to the instructions. The presence of the researcher during the procedure was often helpful in answering questions and providing clarifications. The dissemination of the questionnaires took place in various public places in Athens and the respondents were approached randomly. Caution was taken to preserve an as-random-as-possible sample. For example, the location of the interviews changed frequently. Some background information on the state of the market for Travel Information Systems in the Athens urban area can be found in Polydoropoulou and Tsirimpa (23).

Some general characteristics of the respondents’ responses are presented next. The total number of respondents was 105, each providing responses for 8 choice scenarios, for a total of 840 observations. The gender distribution was balanced (51 male and 54 female respondents). 57% of the respondents (60 out of 105) indicated that they have access to a GPS-based device in their car. The family monthly income distribution that was reported by the respondents is also reasonable: 17% indicated a family monthly income lower than 900€, 12% between 901 and 1300, 19% between 1301 and 1800, 25% between 1801€ and 2300€, 15% between 2301€ and 3000€, 5% between 3001€ and 3800€, 5% between 3801€ and 4500€ and the remaining 2% above 4500€. The data was further post-processed to infer additional variables that could be useful for the analysis. For example, the respondents indicated their residence (at the municipality level). This information was then further aggregated into regions within the Athens metropolitan area, in a process similar to that followed in Polydoropoulou et al. (24). Out of the 105 final respondents, 25 indicated center/downtown as their residence, 51 locations were classified as suburbs (zones in the Northern and Southern parts of the Athens metropolitan area) and all others were clustered together.

Furthermore, age and income were processed to develop the following dummy variables:

- Age up to 25 years old
MODEL ESTIMATION RESULTS

The models have been implemented and estimated in the R language for statistical computing (25), with the pglm (26) package for model estimation and the arm package (27) for parts of the visualization. Table 1 presents the model estimation results for the ordered probit and the random-effects ordered probit models. All models have been specified with the same number of variables, but only those variables that are statistically significant have been retained in the finally estimated models (i.e. after estimating the same models with the variables that were expected to play a role, the models that included some insignificant variables have been re-estimated without these variables). The estimated coefficients for the random-effects ordered probit model are similar to those estimated in the ordered logit model, but in general the t-values are higher. The only exception is the coefficient for the low income, which is still significant at a 90% level and—more importantly—is included in a very significant interaction term with young respondents (less than 25 years old). The estimated coefficient for this variable in the two models has a different sign; however this change is only a small fraction of the value of this interaction term, and therefore the combined effect is in the same direction.

Furthermore, returning to the comparison between the ordered probit and the random-effects ordered probit models, the estimated coefficient for the standard deviation of the random effect is very significant, indicating that indeed there was some heterogeneity in the population that could not be captured by the ordered probit model. From a more global point of view, lower Akaike Information Criterion (AIC) and higher log-likelihood values also indicate that the random-effects ordered probit model is superior to the corresponding ordered probit.

The presented coefficients in the ordered logit model are all significant at a 95% confidence interval, with the exception of the dummy variable for those that have public transit as usual mode, which is significant at a 90% level. The signs and magnitudes of all coefficients are reasonable and consistent with intuitive expectations. The dummy variable for those that are usually driving both in urban and rural roads, which is included in an interaction term with GPS ownership, has been removed from the model as it is statistically insignificant and its exclusion does not affect the values of the estimation results; if it had been retained in the model its coefficient value would have been -0.002 and the t-value equal to -0.019.

In interpreting the coefficients values, one needs to remember that (consistently with common practice) the responses have been re-ordered so that the first option is always the cheaper one, which implies that positive estimated coefficients indicate tendency towards the less expensive alternative. Ceteris paribus, people prefer the cheaper and the more accurate options, while they are driven away from systems that collect more information about them. Those that own GPS devices show a tendency towards the less expensive service; more than half of this effect, however, is countered by the interaction term between GPS ownership and driving in both urban and rural roads. Those that use their GPS device at least occasionally, indicate a tendency towards the more expensive option, potentially indicating that they value the service offered. Male respondents show a preference towards the more expensive option, a phenomenon that seems to be higher for
those that also have kids. Younger people (less than 25 years old) show a significant preference for the more expensive option, possibly reflecting the relatively high disposable income that they may have (as the majority of people in this age group in Greece live with their parents), while people older than 50 years show a preference for the less expensive option, possibly due to a low familiarity with the additional, advanced features that may come with the more expensive options. In the selected, random-effects ordered probit model, those with lower family income show a preference towards the less expensive option, which is an intuitive finding. Those that live in the suburbs show a preference towards the more expensive, and presumably more accurate, option, possibly as they face a longer commute and they have more to gain from a more accurate system. While the parameter for those that use public transport as their usual mode is not significant in the selected random-effects model, its estimated magnitude and sign in the ordered-probit model is interesting. Indeed, it indicates that public transit users show a higher tendency towards the less expensive option, possibly since they are not frequent users of the service and therefore stand to gain less from its use.

Since the respondents provided their preference for each alternative in a seven-level Likert scale, six threshold values between these seven levels can be estimated. In this case these six values are estimated as an intercept and five threshold parameter values. The additional threshold values are obtained as the sum of the intercept and each of these parameter values. For example, the first threshold value is 1.221 (intercept) the 2nd threshold value is 1.221+0.586=1.807 (intercept + mu_1) and the third would be 1.221+1.033=2.254 (intercept + mu_2).

Gender-based market segmentation

The model estimation results in the previous section suggest that gender plays a significant role in the decision to use a service. In this section, separate models are developed for male and female respondents, in order to obtain some insight into the specific factors that differentiate between the two genders (Table 1).

Figure 2 presents the estimated coefficients (and their standard errors) for the gender-based random-effects ordered probit models. For each variable, the values from the model for the female respondents are indicated above those from the model for the male respondents. Furthermore, the values for the female respondents are indicated in red and those for the male respondents in blue. In each case, the estimated coefficient is indicated by the point, the one standard deviation with the thick line and the two standard deviations with the thinner line. The main difference between the two models (besides the coefficients that are only present in one of the models) is in the estimated coefficient for the dummy variable indicating whether somebody drives both in urban and rural roads, which actually has a negative coefficient for the female respondents. However, this variable is not significant (as can also be seen by the fact that its confidence intervals include 0), but it has been maintained, as its interaction with GPS ownership is significant.
### TABLE 1. Random-effects ordered probit model estimation results.

<table>
<thead>
<tr>
<th>Variables</th>
<th>All respondents</th>
<th></th>
<th>Female</th>
<th></th>
<th>Male</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est.</td>
<td>t value</td>
<td>Est.</td>
<td>t value</td>
<td>Est.</td>
<td>t value</td>
</tr>
<tr>
<td>Intercept</td>
<td>1.466</td>
<td>7.18</td>
<td>1.456</td>
<td>5.298</td>
<td>1.052</td>
<td>4.002</td>
</tr>
<tr>
<td>Cost (€/month)</td>
<td>0.209</td>
<td>3.753</td>
<td>0.156</td>
<td>1.998</td>
<td>0.288</td>
<td>3.622</td>
</tr>
<tr>
<td>Accuracy</td>
<td>-0.458</td>
<td>-5.673</td>
<td>-0.41</td>
<td>-3.607</td>
<td>-0.52</td>
<td>-4.524</td>
</tr>
<tr>
<td>Operation w.r.t. privacy level</td>
<td>0.483</td>
<td>3.959</td>
<td>0.394</td>
<td>2.301</td>
<td>0.623</td>
<td>3.557</td>
</tr>
<tr>
<td>Own GPS</td>
<td>0.772</td>
<td>2.51</td>
<td>0.886</td>
<td>2.995</td>
<td>0.573</td>
<td>2.426</td>
</tr>
<tr>
<td>Use GPS at least occasionally</td>
<td>-0.602</td>
<td>-2.595</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drive both in urban and rural roads</td>
<td>0.618</td>
<td>3.34</td>
<td>-0.243</td>
<td>-1.11</td>
<td>1.38</td>
<td>5.993</td>
</tr>
<tr>
<td>Male</td>
<td>-0.521</td>
<td>-3.468</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age less than 25 years</td>
<td>-1.632</td>
<td>-5.483</td>
<td>-1.269</td>
<td>-3.641</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age more than 50 years</td>
<td>0.65</td>
<td>4.186</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Family income less than 1800€/month</td>
<td>-0.193</td>
<td>-1.587</td>
<td>0.45</td>
<td>1.872</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Salaried office worker</td>
<td>-0.455</td>
<td>-2.445</td>
<td></td>
<td></td>
<td>-0.607</td>
<td>-2.727</td>
</tr>
<tr>
<td>Male with children</td>
<td>0.519</td>
<td>4.068</td>
<td>0.364</td>
<td>2.129</td>
<td>0.514</td>
<td>2.81</td>
</tr>
<tr>
<td>Residence: Suburb</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Usual mode: public transport</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Own GPS and drive both in urban and rural roads</td>
<td>-0.85</td>
<td>-3.306</td>
<td>-1.38</td>
<td>-3.938</td>
<td>-1.181</td>
<td>-3.77</td>
</tr>
<tr>
<td>Age less than 25 years and family income less than 1800€/month</td>
<td>2.443</td>
<td>6.529</td>
<td>1.98</td>
<td>4.702</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Threshold parameters for index model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mu_1</td>
<td>0.759</td>
<td>12.76</td>
<td>0.62</td>
<td>8.587</td>
<td>0.928</td>
<td>9.409</td>
</tr>
<tr>
<td>mu_2</td>
<td>1.374</td>
<td>18.882</td>
<td>1.183</td>
<td>12.924</td>
<td>1.599</td>
<td>13.777</td>
</tr>
<tr>
<td>mu_3</td>
<td>1.815</td>
<td>22.651</td>
<td>1.555</td>
<td>15.395</td>
<td>2.136</td>
<td>16.725</td>
</tr>
<tr>
<td>mu_4</td>
<td>2.283</td>
<td>26.24</td>
<td>1.938</td>
<td>17.682</td>
<td>2.732</td>
<td>19.468</td>
</tr>
<tr>
<td>mu_5</td>
<td>3.094</td>
<td>30.356</td>
<td>2.608</td>
<td>20.611</td>
<td>3.807</td>
<td>22.065</td>
</tr>
<tr>
<td><strong>Standard deviation of random effect</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sigma</td>
<td>1.226</td>
<td>15.809</td>
<td>1.214</td>
<td>6.852</td>
<td>1.306</td>
<td>12.449</td>
</tr>
<tr>
<td><strong>Summary statistics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of observations</td>
<td>840</td>
<td></td>
<td>432</td>
<td></td>
<td>408</td>
<td></td>
</tr>
<tr>
<td>Initial log-likelihood</td>
<td>-1627.9</td>
<td></td>
<td>-828.4</td>
<td></td>
<td>-781.8</td>
<td></td>
</tr>
<tr>
<td>Final log-likelihood</td>
<td>-1456.8</td>
<td></td>
<td>-747.5</td>
<td></td>
<td>-683.3</td>
<td></td>
</tr>
<tr>
<td>AIC</td>
<td>2955.5</td>
<td></td>
<td>1528.9</td>
<td></td>
<td>1396.6</td>
<td></td>
</tr>
</tbody>
</table>
FIGURE 2. Comparison of estimated coefficients for gender-based, random effects ordered probit models (female: bottom/red lines, male: top/blue lines, indicating estimate and 2x standard deviations)
**Application**

This section presents an application of the estimated models, in order to demonstrate their use in a practical setting. In particular, the model is run multiple times for different combinations of attributes and their outputs are presented visually and compared. The setup of this experiment keeps most attributes fixed and varies the following attributes among the two alternatives:

- **Cost of service:** the difference between the two alternatives takes values between 0 and 6 Euro/month. This is the same range as the values that were considered in the questionnaire (smaller value of 2 Euro/month and higher value of 8 Euro/month). These values are reasonable and comparable with subscription-based navigation services for mobile devices, e.g., VZ Navigator (Verizon) for the iPhone costs US$4.99/month (in July 2013).

- **Level of privacy:** the same three levels that were considered in the survey are applied: (i) no information collected by the system, (ii) position and destination are collected by the system, and (iii) full trajectory information is collected by the system, allowing the system provider to infer information on driving behavior and potentially driver characteristics.

- **Gender:** male vs. female.

The other attributes (e.g., socioeconomic characteristics) were set at their “average” values, or values corresponding to the majority of the respondents. For example, these application results are based on the assumption that the respondents own a GPS, drive both in urban and rural areas, have a family income between 1800 and 2300 Euro/month, have no children and live in a suburb (according to the definition presented above). Different assumptions would result in a different set of application results; however the overall trends would remain. As it is impossible to show results for all cases, a representative scenario is used for this case.

Figure 3 summarizes the results of this application. The left column presents results for males, while the right column presents the corresponding results for females. The top row presents the results in which both alternative systems require no information from the user (for example, a travel time table is downloaded to the device, where it is used to generate the route for the driver, without a need to transmit any information about the driver to a central facility). The second row assumes that one system requires no information (as above) but the other system operates using limited information about the user, in particular the current location (obtained by the GPS trace) and destination (as indicated in order to generate the route and guidance). The third row assumes that one of the systems again requires no information about the user, but the other obtains detailed information about the movements of the driver (e.g., position, speed and acceleration at regular intervals), thus being able to infer the travel behavior of each user.

The x-axis of each plot represents the difference in travel cost subscription for the two systems. In the cases that have a different privacy level, it is implied that the less expensive system is the one that requires additional information about the user. Assuming a base monthly subscription cost of 2 Euro/month, the range of values of cost difference between 0 and 6 Euro/month results in a range of subscription cost between 2 and 8 Euros/month. This range is reasonable (as evidenced by subscription services currently being offered in the market, such as VZ Navigator, which is available for US$4.99/month) and also the same range as the range of values considered in the survey questionnaire.
The y-axis represents the cumulative probability of choosing between the two alternative services. Since there are seven response levels in the estimated models (and therefore six thresholds between them), there are six curves presented in each plot (and a dashed horizontal line indicating the 100% cumulative probability. In the top row, both services have the same privacy properties and at the right-most edge of the curves, where the cost difference is zero, then the probability of choosing between the two services is pretty much
balanced. As the cost difference increases, then the probability of choosing the more expensive option decreases progressively. Moving down the first column, there is a bigger propensity towards the cheaper option, implying that the probability to give up privacy is higher when the offered service is cheaper. Similar results are obtained for the female respondents in the right column of subfigures. Comparing the magnitude of the choice probabilities of women vs. men, it can be inferred that women are willing to give out less personal information (relative to men).

Based on the estimation results of the models presented in Table 1, the marginal rates of substitution shown in Table 2 can be obtained for the value of privacy and accuracy of the hypothetical monthly service considered in this research. In particular, the average amount that the respondents would like to receive in order to give up one level of privacy is 2.2€/month. Similarly, they would be willing to pay around one €/month (0.96€/month) for a service with a superior accuracy.

<table>
<thead>
<tr>
<th>TABLE 2. Overview of calculated marginal rates of substitution.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Value of privacy</td>
</tr>
<tr>
<td>Value of accuracy</td>
</tr>
<tr>
<td>Ordered logit models</td>
</tr>
<tr>
<td>All respondents</td>
</tr>
<tr>
<td>Random effects ordered logit models</td>
</tr>
<tr>
<td>All respondents</td>
</tr>
<tr>
<td>Female respondents</td>
</tr>
<tr>
<td>Male respondents</td>
</tr>
</tbody>
</table>

The values obtained from the models estimated from the female respondents only show higher values for both of these measures. This indicates that women appear to be more sensitive to privacy, but also show a slightly higher willingness-to-pay for improved accuracy. The findings in the literature about the difference in perception towards privacy of male vs. female respondents are mixed. Cottrill and Thakuriah (28) present a thorough overview of female/gender issues in the context of privacy. Some researchers have indicated that women have a greater tendency to provide information about them (self-disclosure) (29, 30). On the other hand, there is evidence of stronger preference for privacy among women (31, 32). Hann et al. (8) find that online information privacy-related trade-offs did not vary with personal characteristics such as gender.

**Conclusion**

This research presented a framework for the estimation of the willingness to accept giving up privacy in the context of mobile-device based ATIS. Besides investigating the impact of various parameters, separate models have been estimated for female and male respondents. Comparing the magnitude of the choice probabilities of women vs. men, it can be inferred that women are willing to give out less personal information (relative to men). This finding
is confirmed when the marginal rate of substitution between privacy settings and cost of
service is computed: the average amount that the respondents would like to receive in order
to give up one level of privacy is 2.2€/month, while women would demand 2.6€/month and
men 1.8€/month.

One limitation of this research is that it considered three “equidistant” privacy levels: i.e.
the assumption was made that the same “distance” separated the three levels of privacy.
Further studies could design a survey that attempts to distinguish finer differences among
the different levels of each of the dimensions of the experimental design.

ACKNOWLEDGEMENT
The authors thank Mr. Dimitris Veloutsos, student at the School of Rural and Surveying
Engineering of the National Technical University of Athens for help with the data
collection effort.

REFERENCES
emerging data collection technologies and their impact on traffic management
applications. European Transport Research Review, Volume 3, Number 3, 139-148.
Traffic Data Obtained via GPS-Enabled Mobile Phones: the Mobile Century Field
Experiment. Working paper, UCB-ITS-VWP-2009-8, UC Berkeley Center for Future
Urban Transport.
(4) Biem, A., E. Bouillet, H. Feng, A. Riabov, O. Versheure, H.N. Koutsopoulos, M.
Research Record: Journal of the Transportation Research Board, No. 2215,
Transportation Research Board of the National Academies, Washington, D.C., pp.
67-74.
Survey Analysis. Proceedings of the 91st Transportation Research Board Annual
(8) Hann, I.-H., K.-L. Hui, T.S. Lee and I.P.L. Png (2002). Online Information Privacy:
Measuring the Cost-Benefit Trade-Off. Twenty-Third International Conference on
(9) Grossklags, J. and A. Acquisti (2007). When 25 Cents is too much: An Experiment
on Willingness-To-Sell and Willingness-to-Protect Personal Information.
Proceedings of Sixth Workshop on the Economics of Information Security (WEIS
(10) Guderian, E. D. A New Revenue Generating Method for Transportation Funding:
The Vehicle Miles Traveled Fee. In Compendium: Papers on Advanced Surface


