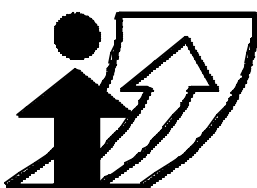


***Estimating non-linear utility functions of time
use in the context of an activity schedule
adaptation model***

**Chang-Hyeon Joh, Theo Arentze and Harry Timmermans
Eindhoven University of Technology**

**Conference paper
Session: Scheduling**



**Moving through nets:
The physical and social dimensions of travel**

10th International Conference on Travel Behaviour Research
Lucerne, 10-15. August 2003

Estimating non-linear utility functions of time use in the context of an activity schedule adaptation model

Chang-Hyeon Joh, Theo Arentze and Harry Timmermans

Eindhoven University of Technology

Urban Planning Group

P.O. Box 513, 5600 MB Eindhoven

The Netherlands

Phone: +31-40-2472283; Fax: +31-40-2475882; e-mail: eirass@bwk.tue.nl

Abstract

Progress in activity-based modeling has recently focused on scheduling and rescheduling decisions (e.g. Gärling *et al.*, 1999). In contributing to this line of research, the authors suggested a comprehensive model, called Aurora, of activity rescheduling decisions as a function of time pressure (Timmermans, *et al.*, 2001). While the original paper focused on duration adjustment and schedule composition, later the proposed theory was elaborated and extended to include many different facets of activity rescheduling behavior (Joh *et al.*, 2001, 2002). Numerical simulations supported the face validity of the model. Given the potential of the model, the next phase in the research project is concerned with the estimation of this complex, non-linear model. This paper develops and tests an appropriate estimation method for the model, using a combination of theory and dedicated genetic algorithms. Results of numerical experiments are discussed. The paper first summarizes the theory underlying the model. Next, a method to estimate the model is proposed. The properties of this method are explored numerically using simulated data. The estimation method is tested using both perfect and noisy data. Finally, some conclusions are drawn and avenues for future research are suggested.

Keywords

Aurora, (re)scheduling, utility function, heuristic estimation, genetic algorithms

Preferred citation

Joh, C.H., T.A. Arentze and Harry J.P. Timmermans (2003) Estimating non-linear utility functions of time use in the context of an activity schedule adaptation model, *Paper presented at the 10th International Conference on Travel Behavior Research*, Lucerne, August 2003.

1. Introduction

Progress in activity-based modeling has recently focused on the analysis and modeling of the process of scheduling and rescheduling activities. An example of such research is Cärling *et al.* (1999), who addressed the problem of how activity scheduling is influenced by time pressure. They suggested that when faced with time pressure, individuals will first try to compress the duration of activities, or try to reschedule them. If this is not sufficient, then individuals are assumed to prioritize activities and eliminate the one with the lowest priority. This process is continued until the total duration is below some threshold.

Over the last few years, the authors developed a more comprehensive model, called Aurora. The model can be viewed as a successor of the Albatross model (Arentze and Timmermans, 2000) and is conducted in the context of the Amadeus research program¹. The system is more comprehensive in that it allows modeling the dynamics of activity scheduling and rescheduling decisions as a function of unexpected events during the execution of activity programs. The model incorporates several behavioral principles and decision styles, including risk-avoiding and opportunistic behavior. While the original paper focused on duration adjustment and schedule composition (Timmermans *et al.*, 2001), later the proposed theory was elaborated and extended to include many different facets of activity rescheduling behavior (Joh *et al.*, 2001a). Several numerical simulations supported the face validity of the model.

Ultimately, however, the model should be derived from empirical data. The question then becomes how the parameters of the model can be estimated. This is certainly not a trivial question as the researcher is faced with several problems. First, the model has no algebraic solution. Secondly, the theory underlying the model argues that scheduling decisions are state-dependent. Finally, the model should satisfy several sets of discontinuous constraints.

¹ Amadeus is a collaborative research program that aims at developing models to assess the long-term, mid-term and short-term implications on activity-travel patterns of multi-modal transportation systems. The program is sponsored by NWO and involves research teams from the universities of Amsterdam, Delft, Eindhoven and Utrecht. The scope and objectives of the program are discussed in Arentze *et al.* (2001).

The present paper reports on the development and test of such an estimation method. First, we will summarize the theory underlying the model. Next, a method to estimate the model is proposed and the properties of this method are explored. Finally, some conclusion will be drawn.

2. Theory

2.1 Activity utility function

Central to the model is that individuals derive some utility from being involved in activities. This utility U is a function of the duration v of the activity. In particular, the functional form can be expressed as:

$$U = \frac{U^{\max}}{(1 + g \exp[\mathbf{b}(\mathbf{a} - v)])^{1/g}} \quad (1)$$

and

$$U^{\max} = \frac{U_x}{1 + \exp[\mathbf{b}_x(\mathbf{a}_x - T)]} \quad (2)$$

where,

\mathbf{a} , \mathbf{b} , g , \mathbf{a}_x , \mathbf{b}_x and U_x are activity-specific parameters, the values of which differ between activities;

v is the duration of the activity;

T denotes the activity history, i.e. the time elapsed since the last implementation of the activity;

U^{\max} asymptotically determines the maximum utility of the activity, which is a function of T .

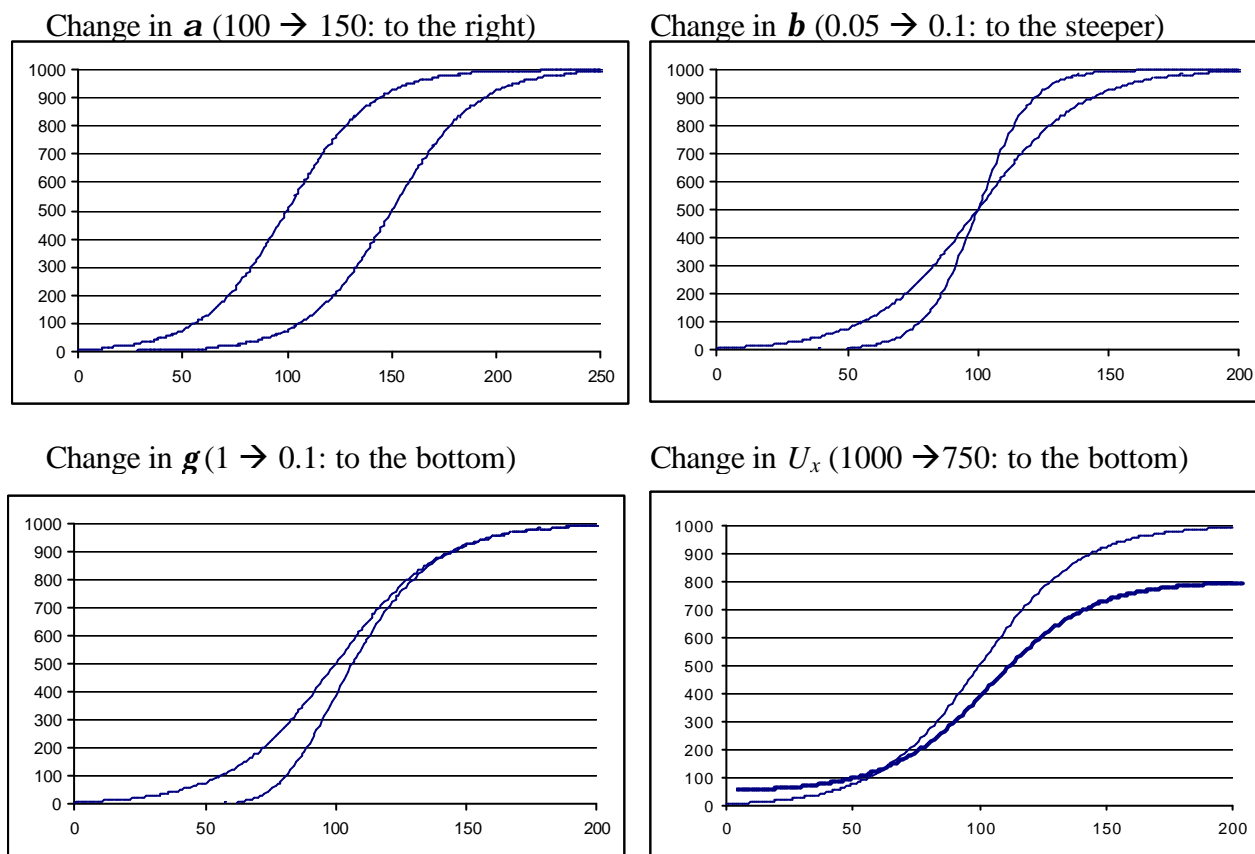
Activity utility is assumed to be a monotonically increasing function of activity duration. Unlike previous authors (e.g., Becker, 1965; Kitamura, 1984), we assume that the utility first increases at an increasing rate and then increases at a decreasing rate. That is, we assume that involvement in an activity first involves an unsaturated, warming-up period where each additional unit of involvement (duration) increases utility at an increasing rate. After reaching the maximum marginal utility at some duration (inflection point), experiences become saturated, and the marginal utility decreases with every additional time unit of further involvement. This notion can be captured in terms of an asymmetric S-shaped curve with an inflection point, as opposed to the commonly assumed logarithmic curve of monotonically decreasing marginal utility.

Figure 1 portrays the model and illustrates the effect of the parameters. The \mathbf{a} parameter determines the duration at which the *marginal* utility reaches its maximum value (inflection point). A larger \mathbf{a} -value shifts the curve to the right, implying a longer warming-up period. The \mathbf{b} parameter determines the slope of the curve. A larger \mathbf{b} -value represents a steeper curve, implying that some fixed change in duration results in a larger change in utility. Thus, a larger \mathbf{b} -value means that the utility is more sensitive to the duration and hence less flexible in terms of adaptation. The \mathbf{g} parameter determines the relative position of the inflection utility. If the value is close to 1, the curve approximates a symmetric curve, and the inflection point is in the middle between the maximum utility and zero. When the value approximates 0, the utility at the inflection level is close to zero, implying that marginal utility is diminishing at virtually all levels of duration.

The maximum utility U^{\max} of an activity is not fixed, but is assumed to be determined by the context of the current schedule. We postulate that U^{\max} depends on the activity history, i.e. the time elapsed since the last implementation of the activity. When an activity has been conducted a long ago, the utility of conducting that activity is assumed to increase. We postulate that the functional form describing the impact of history on the maximum utility is also S-

shaped. The utility of conducting an activity when it just has been performed will be small. The need to conduct that activity again will then systematically increase, until a certain upper limit of U^{\max} has been reached. We simplify the functional form of U^{\max} to the symmetric case.

Figure 1. Impacts of utility parameters



The utility of an activity can then be computed, given the estimated a , b , g , a_x , b_x and U_x , and the duration v and history T of the activity. Note that the model can be extended by further incorporating into the U^{\max} function other decision dimensions such as activity sequence, location, transport mode, accompanying person, etc.

2.2 Schedule adjustment

Faced with time pressure, an individual is assumed to proceed with a series of mental processes of schedule adjustment and implement the best schedule. In the proposed model, various operators are assumed to accomplish schedule adjustments (Joh *et al.*, 2002). The schedule composition operator changes the list of activities included in a schedule by deleting, inserting and substituting activities. Sequencing, location, transport mode and accompanying person operators change the means of implementing the concerned activity. The application of the operators leads to incremental mental adjustments of the schedule and continues until no more improvement is possible.

2.3 Problems with estimation

Assume that we know in advance the maximum level of utility U^{\max} of each activity, and we can observe the level of utility U of each activity over time. We could then linearize the suggested activity utility function and, using the information of duration v of the observed schedule data, directly solve the equation for the parameter values α , β and γ for each activity. This is unfortunately not the case because the utilities of the activities (U and U^{\max}) are unobservable and unknown. Furthermore, unlike the logarithmic function of ever-diminishing marginality (e.g., Kitamura, 1984), the suggested utility function of the current research does not provide the algebraic solution for the optimum durations of the maximum schedule utility. Therefore, we cannot directly 'solve' the estimates but have to 'find' the estimates that best fit the observed durations by searching iteratively multiple combinations of parameter values. The number of such combinations would however be prohibitively large for an exhaustive search.

2.4 Estimation method

For the above reasons, we suggest a heuristic method that searches only part of the entire solution space, but at the same time, provides good, near optimum solutions. The method is based on the critical assumption that the marginal utility of activities in the schedule is the same. If this would not be the case, an individual would further adjust durations to increase the total utility. The duration of the activity of the higher marginal utility will be increased while the duration of the other activity will be decreased.

In principle, based on these theoretical decisions, an estimation method could be developed. However, the numerical value of the equal marginal utility of the schedule is unknown in reality, and multiple solutions for the parameter estimates that satisfy the equal marginal utility across activities of the schedule could exist. Therefore, an additional assumption is required. To find a solution, we postulate that the level of equal marginal utility of the schedule is partly reflected by the amount of time pressure of the schedule, measured as the total duration of the *fixed* activities of the schedule, and the number of activities. The numerical level of equal marginal utility can then be approximated as a function of these variables. The rationale behind this assumption is that given the number of activities, higher time pressure (less available time) would raise the level of equal marginal utility. On the other hand, given the time pressure, a larger number of activities would raise the marginal utility for each activity. These assumptions seem appropriate given our theory especially for saturated activities. Equation (3) expresses this critical assumption.

$$\frac{bU^{\max} \exp[b(a - v)]}{(1 + g \exp[b(a - v)])^{1/g+1}} = \sum d_k X_k \quad \forall a \in S \quad (3)$$

where,

X_k is the k^{th} attribute of the equal-marginal-utility function ($X_1 =$ fixed duration, $X_2 =$ number of activities of the schedule);

d_k is the marginal contribution of the k^{th} attribute X_k to the level of equal marginal utility, which is attribute-specific;

S denotes the current schedule.

By solving equation (3) for v , given each of the combinations of possible values of parameters, the overall goodness-of-fit of the predicted set of parameter values can be calculated as:

$$G_l = \sum G_{a_i} \quad (4)$$

with

$$G_{a_i} = |v_a^o - v_{a_i}^p| \quad (5)$$

where,

G_l is the goodness-of-fit of the l^{th} predicted combination of parameter estimates, where $l = 1, \dots, L$, and L is the number of the parameter values combinations to be examined;

G_{a_i} is the goodness-of-fit of the l^{th} predicted combination of parameter estimates for activity a ;

v_a^o and $v_{a_i}^p$ are respectively the observed duration of activity a and the duration of activity a predicted by the l^{th} predicted combination of parameter values.

The set of parameter values that corresponds to the best goodness-of-fit can then be identified.

Still, however, some further operational problems need to be solved for obtaining the duration prediction v^p in order to compute the goodness-of-fit of the associated set of predicted parameter estimates. First, there is no direct, algebraic solution for v^p that satisfies equation (3), and therefore, we used an algorithm of golden section search to ‘find’ the duration instead of ‘solving’ it.

Secondly, the predicted set of parameter values may have no cross points between the equal marginal utility line given by the RHS of equation (3) and the marginal utility curve given by

the LHS of equation (3). That is, it may be the case that the predicted time pressure goes beyond the maximum of the marginal utility. It then has no cross points for the corresponding duration prediction. In reality, it may be that when the time pressure is so high, and the schedule requires very high marginal utilities for activities to be included, the activities with a lower maximum marginal utility will not survive, and therefore, we would not observe such activities in the schedule. In our estimation, however, we do observe the activity that was implemented in the schedule, and therefore, if the predicted parameters result in no cross points, that predicted solution would clearly be wrong. The predicted parameters associated with such ‘illegality’ and their neighbors should not be visited again in the iterative solution search procedure enforced by assigning an appropriate size of penalty. To this end, the goodness-of-fit of an activity should be revised from equation (5) as:

$$G_{a_i} = \begin{cases} |v_a^o - v_{a_i}^p| & \text{for legal solution} \\ |v_a^o - 0| + 1440 + gap & \text{for illegal solution} \end{cases} \quad (6)$$

As implied by this equation, the measure of the goodness-of-fit of the illegal solution for an activity is the sum of three sources of differences. The first source of difference is the pure difference between observation and prediction. As discussed above, if the time pressure is much too high compared with the maximum level of an activity’s marginal utility, that activity should not be included in the schedule, and therefore, the predicted duration v^p of this activity is zero. The second source of difference is a penalty for being illegal. In the current study, we used 1 minute as the unit of measurement for duration. The added number 1440 then means the entire time duration of a day in minutes. Equation (6) distinguishes illegal solutions from legal ones by adding this big number, which makes the resulting measure of goodness-of-fit worse than any possible legal solution. The final source of difference for the illegal solutions is the degree of illegality to make the search sensitive for direction. Figure 2 illustrates three different solutions of an activity, where the predicted parameter values of this activity’s marginal utility curve \mathbf{a} , \mathbf{b} , \mathbf{g} , \mathbf{a}_x , \mathbf{b}_x and U_x are the same, and only the time pressure parameter values \mathbf{d} are different. The first solution is legal in the sense that the marginal utility curve and the time pressure line (TP₁) crosses at least one point. The second and third solutions are however illegal, because the marginal utility curve and time pressure lines do not cross. The lower part of equation (6) should be applied. The first two terms of the RHS of the lower part of this equation are the same for the second and third solutions, which makes their goodness-of-fit measure much worse than the first solution. The vertical arrows between the

time pressure lines and the maximum of the marginal utility curve makes a further distinction between these illegal solutions, which states that the third solution is even worse than the second.

Figure 2. Illegality of the solutions

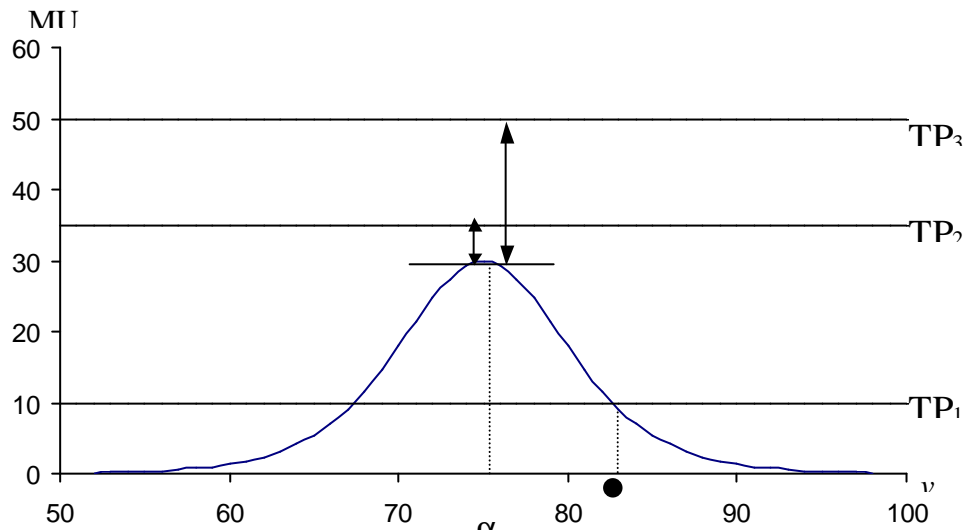
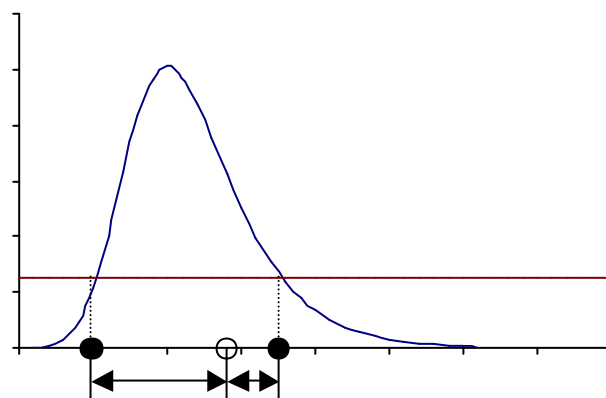


Figure 3. Decision of predicted duration



Finally, the existence of the equilibrium durations at both sides of the inflection point raises the question of how to choose the one that is used for prediction. We let the method choose the one that gives the better match. This can be illustrated as in Figure 3.

In the figure, the black dots are the predicted durations v^p corresponding to the cross points of the observed level of equal marginal utility and the marginal utility curve. The white dot is the observed duration v^o of this activity. The black dot on the RHS is chosen in this illustration because it is closer to the observation than the other black dot. Accordingly, the goodness-of-fit of an activity is again revised from equation (6) as:

$$G_{a_i} = \begin{cases} \max[|v_a^o - v_{a_i}^p|] & \text{for legal solution} \\ |v_a^o - 0| + 1440 + gap & \text{for illegal solution} \end{cases} \quad (7)$$

The proposed estimation method can therefore be summarized as:

$$\text{Minimize: } G \in \{G_l\} \quad (8)$$

$$\text{Subject to: } MU_{a_i} = MU_{a_i} \quad \forall a, a' \in S \quad (9)$$

where,

G is the goodness-of-fit of the estimated model of the best parameter values combination;

G_l is defined as equations (4) and (7);

MU_a denotes the marginal utility of activity a , as expressed in the LHS of equation (3).

Solving equation (3) for v across activities of the schedule meets the constraints of equation (9) and provides the prediction of activity duration used in equation (7). For each activity of the schedule, then the prediction error is computed, and the overall goodness-of-fit of a predicted combination of parameter values is the sum of prediction errors across activities as in equation (4).

3. Algorithm

Our problem thus is to estimate from schedule data the activity-specific parameters \mathbf{a} , \mathbf{b} , \mathbf{g} , \mathbf{a}_x , \mathbf{b}_x and U_x , given the information of duration v , history T and time pressure attribute X . A genetic algorithm was applied to solve equations (8) and (9). The following operational decisions were made.

3.1 Representation of the solution candidates

We employed a real coding scheme to represent the solution candidates of the real parameter values. The real-coding genetic algorithm (RCGA) does not binary-digitalize the real number information but uses real numbers directly representing the solutions with minimum and maximum possible values (Wright, 1991). Given m parameters to estimate for each of n activities and p time-pressure parameters, a solution candidate is represented as an array of $mn+p$ elements of real numbers, and an element of a real number represents a corresponding parameter. In a preliminary study, this RCGA representation scheme outperformed the ordinary binary representation scheme in terms of precision and the speed. In particular, the increased speed was obtained by the RCGA representation scheme, where the encoded information (genotype) for genetic modification and the real form (phenotype) for candidate evaluation are the same, and hence, a decoding process that transforms the genotype information into the phenotype is not required.

3.2 Genetic operators

Our RCGA also employed crossover and mutation operators for genetic modification of the solution candidates like any other GA. The details of the operators of the RCGA are however different from ordinary binary GAs. After a series of preliminary studies, we chose the BLX-0.5 crossover and Mühlhoben mutation (Herrera *et al.*, 1998). As for crossover, assume two solution candidate arrays $C_1 = [c_1^1, \dots, c_{mn+p}^1]$ and $C_2 = [c_1^2, \dots, c_{mn+p}^2]$ that are selected from the current pool of solution candidates to be modified for the next pool. The h_i of the offspring $H = [h_1, \dots, h_{mn+p}]$ is determined such that the value randomly lies in-between $c_{\min} - I \cdot \omega$ and $c_{\max} + I \cdot \omega$, where $1 \leq i \leq mn+p$, $c_{\min} = \min[c_i^1, c_i^2]$, $c_{\max} = \max[c_i^1, c_i^2]$ and $I = c_{\max} - c_{\min}$. We chose 0.5 for ω . As for mutation, assume a randomly selected solution candidate $C = [c_1, \dots, c_{mn+p}]$. The c_i' of the offspring $C' = [c_1', \dots, c_{mn+p}']$ is determined such that $c_i' = c_i \pm 0.1 \cdot (b_i - a_i) \cdot \sum_{k=0}^{15} h_k 2^{-k}$, where a_i and b_i are the minimum and maximum values that the i^{th} parameter can take, h_k is randomly determined to be 1 with a probability of 1/16 and 0 with a probability of 15/16, and the + or – sign is chosen with a probability of 0.5.

3.3 Genetic parameters

After an extensive study, we chose the following operational parameter values for the genetic procedure.

- Size of the pool or Number of solution candidates for an iteration = 100
- Stop condition = 10000 iterations after the initialization
- Probability of choosing crossover instead of mutation for the current round iteration = 70 %
- Number of solution candidates selected for crossover from the previous pool = 50
- Number of solution candidates selected for mutation from the previous pool = 90
- Selection of solution candidates for modification = Random selection with replacement
- Probability of mutating a parameter of the solution candidate selected for mutation = 10 %

3.4 Overall procedure

Overall, our RCGA works as follows. ① The RCGA randomly initializes 100 solution candidates, each of which is an array of $mn+p$ real numbers that represent m parameters of n activities and p time pressure parameters. ② It then evaluates each candidate, which is a rather complex procedure. Each solution candidate is a prediction of parameter values of activity utility function at the current step of iteration. Given these values and the time pressure information of the observed schedules, the system finds the predicted duration of each activity included in the observed schedule using the relation of equality between the marginal utilities mathematically derived from the utility function and observed from the time pressure attributes X . The absolute difference between predicted and observed durations of that activity is then computed, and these differences are summed across activities of the schedule and finally across schedules of the entire data as a measure of goodness-of-fit of the solution candidate. ③ The RCGA selects solution candidates, and the selection probability is proportional to the goodness-of-fit. Better goodness-of-fit increases the chance for a candidate to be selected. The following probabilistic roulette wheel (Joh *et al.*, 2001b) is used at each time of a selection.

$$P_l = \frac{\sum_l G_l}{G_l} \bigg/ \frac{\sum_l \sum_l G_l}{\sum_l G_l} \quad (10)$$

where P_l and G_l are respectively the selection probability and the goodness-of-fit measure of the l^{th} solution candidate ($l = 1, \dots, 100$). Note that the goodness-of-fit of a candidate is the sum of prediction errors across schedules, and hence, a smaller value means a better fit. The selection is repeated 100 times with replacement. ④ Among the selected candidates, the RCGA modifies randomly selected 50 (for crossover) or 90 (for mutation) candidates. ⑤ The RCGA again evaluates the new pool of genetically modified solution candidates. An iteration consists of the steps ③-④-⑤ and is repeated 10000 times for a run.

4. Model Test Results

Before dealing with any empirical data, we examined the properties of the suggested estimation method on simulated data. The purpose of this study was to investigate whether the suggested method produced the correct results using a set of simulated schedule data. Furthermore, we wanted to better understand the performance of the suggested approach for noisy data. To this end, we prepared a set of simulated activity schedule data, which assumed no noise for activity duration and time pressure attributes, and a set of activity schedule data, using the same value with some added amount of noise. In the following, we first examine whether the suggested estimation method is capable of reproducing the parameters, and then further examine the robustness of the suggested method against various sources of noise in the simulated data.

4.1 Estimation results for simulated activity schedules using exact data

The specification of the model that was used for the simulations is largely the same as the one expressed in equations (8) and (9). We however simulated five time pressure attributes, instead of two, and hence we have five time pressure parameters, \mathbf{d}_1 to \mathbf{d}_5 for the purpose of model test.

We assumed two types of activities, which suggests a total of seventeen activity utility parameters ($= 2 \times 6 + 5$) to estimate. The simulated data consists of fifty simulated schedules of these two types of activities. A schedule provides information about the duration v and history T of each activity and the numerical values of five time pressure attributes X of that schedule. The observed history is simulated in the range from 1 to 30 integer values, and the observed time pressure attributes are observed in the range from 0 to 100 real values across cases. The 'true' values for parameters were prescribed arbitrarily but reasonably representing assumed activities. Given these simulated observations and the true values of the activity utility function parameters (Table 1), the simulated duration observations for the activities of each schedule can be obtained from equation (3). As mentioned in section 3.1, the RCGA needs to predefine the minimum and maximum possible values to simulate. The following ranges for the parameter values were used: $\mathbf{a} = 10 \sim 300$; $\mathbf{b} = 0.01 \sim 1$; $\mathbf{g} = 0.1 \sim 1$; $\mathbf{a}_x = 1 \sim 30$; $\mathbf{b}_x = 0.01 \sim 1$; $U_x = 50 \sim 500$.

The proposed estimation method was run 30 times on the same data. Each run was terminated after 10000 iterations. The total of 30 runs took approximately 22 hours to complete, indicating the complexity of the problem. Table 1 shows the estimation results averaged across the 30 runs.

Table 1. Estimation results with exact data

parameter		true value	estimated
a	activity 1	75	75.31
	activity 2	420	421.39
b	activity 1	0.15	0.15
	activity 2	0.10	0.11
g	activity 1	0.8	0.58
	activity 2	0.1	0.27
a_x	activity 1	7	7.64
	activity 2	3	3.06
b_x	activity 1	0.15	0.14
	activity 2	0.20	0.19
U_x	activity 1	250	307.20
	activity 2	150	173.02
d₁		0.0125	0.01587
d₂		0.0500	0.06253
d₃		0.0240	0.02952
d₄		0.0800	0.09671
d₅		0.0100	0.01213
GOF		-	0.09946

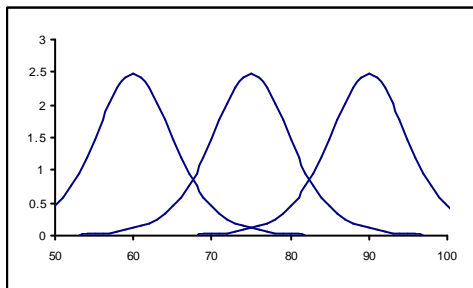
Table 1 suggests that the parameter estimates are close to the true parameter values that were used for generating the simulated schedule data. This means that if our assumption that individual activity rescheduling behavior is based on equalizing marginal utilities is true, and data do not exhibit any noise, the suggested estimation method will produce fairly exact estimates of activity utility parameters.

A more detailed inspection of the results suggest that the **a**, **b**, **a_x** and **b_x** values are estimated more precisely than **g** and **U_x**. This can be explained by examining Figure 4 which portrays the marginal impacts of the parameters and shows that the impact of **g** and **U_x** is relatively small. In case of **g**, it is difficult to recognize the difference in the curve between different **g**'s

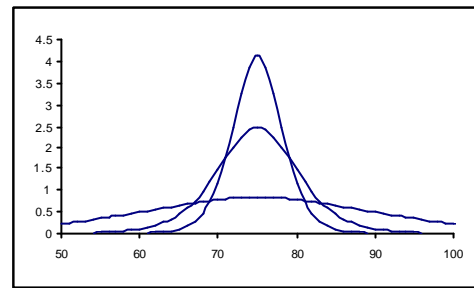
under a certain level of time pressure. In case of U_x , unless the difference is very large (500 and 100), the curves show little difference between U_x 's (500 and 450). The GA is therefore less sensitive to these two parameters in the estimation. In contrast, changes in duration are most sensitive to the a parameter, which is consistent with the estimation results where the a value has the highest accuracy.

Figure 4. Impacts of parameter values on the marginal utility curve

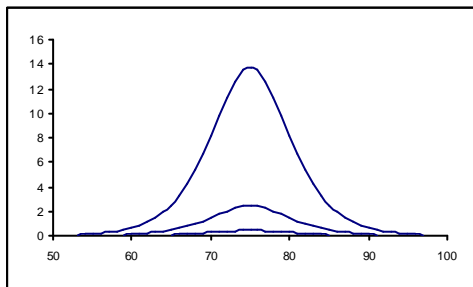
a (60 \rightarrow 75 \rightarrow 90: to the right)



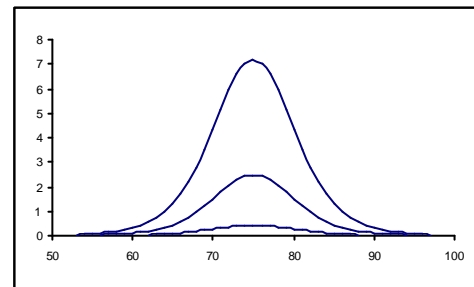
b (0.5 \rightarrow 0.3 \rightarrow 0.1: to the bottom)



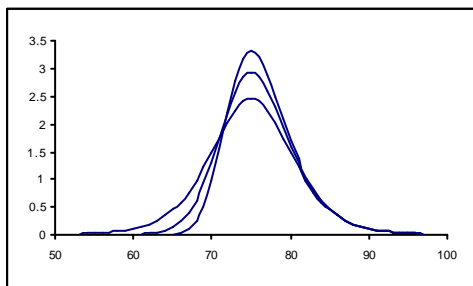
a_x (10 \rightarrow 20 \rightarrow 25: to the bottom)



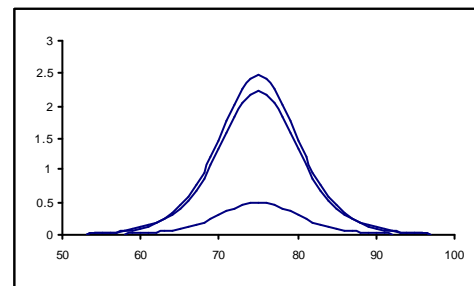
b_x (0.05 \rightarrow 0.1 \rightarrow 0.15: to the bottom)



g (0.2 \rightarrow 0.5 \rightarrow 1: to the bottom)



U_x (500 \rightarrow 450 \rightarrow 100: to the bottom)



An additional concern when estimating these types of functions is the possible linear correlation between parameter estimates. Table 2 shows the correlation between parameter estimates over 30 runs of estimation.

Table 2. Correlations between parameter estimates

Activity 1						
	<i>a</i>	<i>b</i>	<i>g</i>	<i>a_x</i>	<i>b_x</i>	<i>U_x</i>
<i>a</i>	1	-.616**	-.842**	-.415*	.493*	-.567**
<i>b</i>	-.616**	1	.698**	.935**	-.943**	.899**
<i>g</i>	-.842**	.698**	1	.410	-.453*	.654**
<i>a_x</i>	-.415*	.935**	.410*	1	-.990**	.839**
<i>b_x</i>	.493*	-.943**	-.453*	-.990**	1	-.842**
<i>U_x</i>	-.567**	.899**	.654**	.839**	-.842**	1

Activity 2						
	<i>a</i>	<i>b</i>	<i>g</i>	<i>a_x</i>	<i>b_x</i>	<i>U_x</i>
<i>a</i>	1	.610**	-.551**	-.201	.200	-.871**
<i>b</i>	.610**	1	.322	.106	-.097	-.258
<i>g</i>	-.551**	.322	1	.340	-.341	.769**
<i>a_x</i>	-.201	.106	.340	1	-.921**	.402
<i>b_x</i>	.200	-.097	-.341	-.921**	1	-.424*
<i>U_x</i>	-.871**	-.258	.769**	.402	-.424*	1

Note: * and ** denotes that the correlation is significant at the 0.05 and 0.01 levels, respectively.

Table 2 shows that there exist significant linear correlations between parameter estimates. These results suggest that when dealing with the empirical data, a sequential estimation strategy may be preferable if we face substantial interaction between parameter estimates, resulting in unstable parameter estimates over runs. As shown in Figure 4, the *a* parameter is most sensitive to the change in the duration and is thus expected to be most accurate and stable in prediction, which is proven in Table 1. The sequential estimation may therefore begin with the estimation of *a* given other parameter values of averages of the initial simultaneous estimations. In fact, preliminary tests suggest that indeed considerably more stable parameters were obtained when using a sequential estimation procedure. A detailed discussion of these results is beyond the scope of the present paper.

4.2 Estimation results for simulated activity schedules using noisy data

Three types of noise were considered: (i) rounding of reported activity duration, (ii) inexact observation of the time pressure attributes and (iii) inexact observation of activity duration. These sources of noise can be expressed in a single equation as:

$$\frac{bU_x \exp[-\exp[b_x(\mathbf{a}_x - T)]] \exp[b(\mathbf{a} - (\hat{v} + e_{vn}))]}{(1 + g \exp[b(\mathbf{a} - (\hat{v} + e_{vn}))])^{1/g+1}} = \Sigma dX + e_n \quad (11)$$

where,

\hat{v} is the duration rounded by respondents ;

e_n is the error term of the time pressure for the n th observed schedule;

e_{vn} is the error term of the duration observation.

The goal of the analysis here was to examine the robustness of the suggested estimation method in the presence of such noise for each source separately, one at a time.

4.2.1 Rounding noise

The simulated data represent activity duration to the precision of three digits after decimal points. Given the simulated observed T and X , a search algorithm finds the value of v based on equation (3) at this level of precision. Table 3 presents an example of the simulated data, whereas Table 4 presents the estimation results for different degrees of rounding error.

Table 3. Example of the simulated schedule data (no rounding, 1-minute and 5-minutes rounding)

no rounding									
case	v1	v2	T1	T2	X1	X2	X3	X4	X5
1	85.616	422.797	24	25	2.62	29.41	59.2	23.3	23.68
2	96.986	421.676	22	1	9.40	5.72	14.58	2.44	21.95
3	88.115	421.881	17	7	44.35	0.35	45.41	12.54	63.88
4	89.351	427.553	18	8	11.11	32.59	0.34	12.14	19.60
5	83.345	427.412	12	23	40.48	9.26	50.68	14.46	86.35
6	92.574	436.768	26	22	8.27	10.79	22.74	4.25	70.86
7	93.965	438.813	23	16	1.80	5.78	39.95	3.80	22.22
8	75.334	444.023	2	22	21.04	7.56	9.96	0.91	24.98
9	91.649	434.218	27	15	4.65	4.73	34.40	6.20	92.67
10	82.997	425.829	12	17	53.92	17.02	11.46	24.87	54.58
...
1-minute rounding									
case	v1	v2	T1	T2	X1	X2	X3	X4	X5
1	86	423	24	25	2.62	29.41	59.2	23.3	23.68
2	97	422	22	1	9.4	5.72	14.58	2.44	21.95
3	88	422	17	7	44.35	0.35	45.41	12.54	63.88
4	89	428	18	8	11.11	32.59	0.34	12.14	19.6
5	83	427	12	23	40.48	9.26	50.68	14.46	86.35
6	93	437	26	22	8.27	10.79	22.74	4.25	70.86
7	94	439	23	16	1.8	5.78	39.95	3.8	22.22
8	75	444	2	22	21.04	7.56	9.96	0.91	24.98
9	92	434	27	15	4.65	4.73	34.4	6.2	92.67
10	83	426	12	17	53.92	17.02	11.46	24.87	54.58
...
10-minutes rounding									
case	v1	v2	T1	T2	X1	X2	X3	X4	X5
1	90	420	24	25	2.62	29.41	59.2	23.3	23.68
2	100	420	22	1	9.4	5.72	14.58	2.44	21.95
3	90	420	17	7	44.35	0.35	45.41	12.54	63.88
4	90	430	18	8	11.11	32.59	0.34	12.14	19.6
5	80	430	12	23	40.48	9.26	50.68	14.46	86.35
6	90	440	26	22	8.27	10.79	22.74	4.25	70.86
7	90	440	23	16	1.8	5.78	39.95	3.8	22.22
8	80	440	2	22	21.04	7.56	9.96	0.91	24.98
9	90	430	27	15	4.65	4.73	34.4	6.2	92.67
10	80	430	12	17	53.92	17.02	11.46	24.87	54.58
...

As expected, more rounding error results in less accurate estimates, in particular the g and U_x estimates as discussed in Figure 4. Considering the homogeneous activity durations resulting from the rounding, however, the results are rather accurate and promising.

Table 4. Estimation results with rounded data

parameter		true value	no rounding	1 minute rounding	10 minutes rounding
a	activity 1	75	75.31	75.89	80.11
	activity 2	420	421.39	421.80	422.92
b	activity 1	0.15	0.15	0.14	0.18
	activity 2	0.10	0.11	0.11	0.13
g	activity 1	0.8	0.58	0.27	0.76
	activity 2	0.1	0.27	0.48	0.87
a_x	activity 1	7	7.64	6.61	9.14
	activity 2	3	3.06	3.02	3.18
b_x	activity 1	0.15	0.14	0.15	0.13
	activity 2	0.20	0.19	0.20	0.19
U_x	activity 1	250	307.20	248.52	171.89
	activity 2	150	173.02	176.77	142.70
d₁		0.0125	0.01587	0.01603	0.00979
d₂		0.0500	0.06253	0.06445	0.04069
d₃		0.0240	0.02952	0.03003	0.02402
d₄		0.0800	0.09671	0.09735	0.07987
d₅		0.0100	0.01213	0.01211	0.01197
GOF		-	0.09946	0.34374	11.81063

4.2.2 Time pressure attribute observation noise

To study this effect as expressed as e_n in the RHS of equation (11), two scales of the error were introduced using a normal distribution $N(0, \sigma^2)$. One has a standard deviation size of 10 % of the average time pressure level 2.7, and the other the standard normal distribution. That is, $e_n \sim N(0, 0.27^2)$ and $e_n \sim N(0, 1)$. The results are described in Table 5.

Several interesting observations can be made from Table 5. First, the estimates appear to be relatively robust against this type of noise. Secondly, the scale seems to affect the results. If the estimation method is perfect, the introduction of a random error term e_n in the RHS of equation (11) should not affect the results. Given the limited number of simulated observations and the small size of time pressure, however, the result is not very disappointing. The random error of scale $\text{std} = 1$ seems to confuse the GA too much. Increasing the number of observations should improve the estimation results. Optionally, a sequential estimation of the time pressure parameters would improve the accuracy, which takes a linear regression analy-

sis of equation (3) with the scalar value of estimated marginal utility in the LHS of the equation.

Table 5. Estimation results with noisy time-pressure data

parameter		true value	no noise	std = 0.27	std = 1
a	activity 1	75	75.31	77.93	80.37
	activity 2	420	421.39	424.66	426.76
b	activity 1	0.15	0.15	0.17	0.19
	activity 2	0.10	0.11	0.13	0.13
g	activity 1	0.8	0.58	0.58	0.16
	activity 2	0.1	0.27	0.19	0.18
a_x	activity 1	7	7.64	6.93	11.16
	activity 2	3	3.06	3.42	3.71
b_x	activity 1	0.15	0.14	0.16	0.12
	activity 2	0.20	0.19	0.16	0.27
U_x	activity 1	250	307.20	211.23	361.45
	activity 2	150	173.02	126.98	178.28
d₁		0.0125	0.01587	0.01767	0.03461
d₂		0.0500	0.06253	0.06288	0.09891
d₃		0.0240	0.02952	0.03223	0.07479
d₄		0.0800	0.09671	0.09972	0.09974
d₅		0.0100	0.01213	0.00759	0.02435
GOF		-	0.09946	4.59547	11.60373

4.2.3 Duration measurement noise

Measurement error e_{vn} is added to the duration both in the numerator and denominator of the LHS of equation (11) as $v + e_{vn}$. Errors were assumed to be drawn at random from a normal distribution $N(0, \mathbf{s}^2)$. Table 6 gives an example, where “std10% act1” means that the error is randomly drawn from a normal distribution of size \mathbf{s} , which is 10 % of the average duration of activity 1 in the data. The average simulated duration of activity 1 and activity 2 is 86 and 429 minutes, respectively. Hence the “std10%” condition includes the error $e_{vn} \sim N(0, 8.6^2)$ for activity 1 and $e_{vn} \sim N(0, 42.9^2)$ for activity 2 added to the duration v in both the numerator and the denominator of the LHS of equation. The “std5%” condition results in $e_{vn} \sim N(0, 4.3^2)$ for activity 1 and $e_{vn} \sim N(0, 21.45^2)$ for activity 2. Table 6 gives an example of the measurement noise simulated in this way.

Table 6. Data with measurement error of different sizes

error size of 10% of average duration						
case	original duration		introduced error		duration with error	
	activity 1	activity 2	std10% act1	std10% act2	activity 1	activity 2
1	83.720	442.837	-16.19	-13.34	67.530	429.497
2	84.856	423.586	-2.76	-76.64	82.096	346.946
3	93.901	436.775	-20.59	31.91	73.311	468.685
4	85.396	431.398	-6.86	-19.63	78.536	411.768
5	91.466	434.846	2.93	-10.68	94.396	424.166
6	90.889	430.142	10.94	-58.65	101.829	371.492
7	87.401	427.610	4.53	39.37	91.931	466.980
8	92.693	422.688	8.47	-41.10	101.163	381.588
9	90.541	430.078	4.88	28.22	95.421	458.298
10	89.534	431.238	-4.73	-1.59	84.804	429.648
...
error size of 5% of average duration						
case	original duration		introduced error		duration with error	
	activity 1	activity 2	std5% act1	std5% act2	activity 1	activity 2
1	83.720	442.837	-8.09	-6.67	75.630	436.167
2	84.856	423.586	-1.38	-38.32	83.476	385.266
3	93.901	436.775	-10.3	15.95	83.601	452.725
4	85.396	431.398	-3.43	-9.82	81.966	421.578
5	91.466	434.846	1.47	-5.34	92.936	429.506
6	90.889	430.142	5.47	-29.32	96.359	400.822
7	87.401	427.610	2.27	19.68	89.671	447.290
8	92.693	422.688	4.23	-20.55	96.923	402.138
9	90.541	430.078	2.44	14.11	92.981	444.188
10	89.534	431.238	-2.37	-0.79	87.164	430.448
...

To investigate the effect of the duration measurement noise, we tested two sets of simulated schedule data of different sizes. The estimation results are shown in Table 7.

The following observations can be made. As for the sample size, the bigger sample ($n=200$) returns a better result than the smaller sample ($n=50$). As for the goodness-of-fit, the size of the GOF of the estimated model is almost the same with the size of the error introduced to the duration measurement, and the estimations are all converged. The figures of Table 8 are the results averaged across n cases.

Table 7. Estimation results with measurement noise on the data of different sizes

50-cases data					
parameter	true value	no error	std = 10%	std = 5%	
a	activity 1	75	75.31	83.20	82.74
	activity 2	420	421.39	410.46	418.40
b	activity 1	0.15	0.15	0.47	0.24
	activity 2	0.10	0.11	0.04	0.08
g	activity 1	0.8	0.58	0.64	0.64
	activity 2	0.1	0.27	0.53	0.62
a_x	activity 1	7	7.64	17.38	11.09
	activity 2	3	3.06	6.47	6.45
b_x	activity 1	0.15	0.14	0.22	0.32
	activity 2	0.20	0.19	0.34	0.33
U_x	activity 1	250	307.20	207.95	166.70
	activity 2	150	173.02	430.42	390.40
d₁		0.0125	0.01587	0.04447	0.04911
d₂		0.0500	0.06253	0.01751	0.04327
d₃		0.0240	0.02952	0.05218	0.06489
d₄		0.0800	0.09671	0.05001	0.05607
d₅		0.0100	0.01213	0.00807	0.01505
GOF	-	0.09946	44.00144	22.10836	
200-cases data					
parameter	true value	no error	std = 10%	std = 5%	
a	activity 1	75	75.57	79.03	78.58
	activity 2	420	421.31	421.55	423.48
b	activity 1	0.15	0.15	0.18	0.17
	activity 2	0.10	0.11	0.05	0.06
g	activity 1	0.8	0.51	0.70	0.63
	activity 2	0.1	0.44	0.55	0.62
a_x	activity 1	7	7.65	8.91	7.42
	activity 2	3	3.02	6.13	6.05
b_x	activity 1	0.15	0.14	0.14	0.15
	activity 2	0.20	0.19	0.30	0.29
U_x	activity 1	250	288.17	288.00	279.12
	activity 2	150	174.95	326.01	271.48
d₁		0.0125	0.01522	0.02678	0.02406
d₂		0.0500	0.06057	0.06283	0.07522
d₃		0.0240	0.02835	0.03579	0.04032
d₄		0.0800	0.09343	0.09786	0.09879
d₅		0.0100	0.01176	0.01231	0.01332
GOF	-	0.17417	36.42288	18.13449	

Table 8. Goodness-of-fit across error sizes

sample size	error size	GOF	error sum
$n = 50$	std=10%	44.00	47.97
	std=5%	22.11	23.98
$n = 200$	std=10%	36.42	37.41
	std=5%	18.13	18.70

Note: The introduced error sum = $|e_{1n}|+|e_{2n}|$.

As for the accuracy of the estimates, the estimates are not very accurate for some parameters for noisy data. Moreover, the inaccuracy is amplified with the size of the introduced error across parameters of both activities in both data sets. The reason may be the following. When measurement error e_{vn} is introduced from a normal distribution, duration for the actual estimation is changed to $v + e_{vn}$. If the estimation method is insensitive to the introduced random duration error, the marginal utility (the LHS of equation) should be averaged to the true values. In other words, given the marginal utility $MU = f(v)$, $f(v) \approx (f(v+e) + f(v-e))/2$. For example, the marginal utility of the true duration 440 should (more or less) be the same as the average of two marginal utilities of duration 440+40 and duration 440-40. Assume that \mathbf{a} , \mathbf{b} , \mathbf{g} , \mathbf{a}_x , \mathbf{b}_x and U_x are 75, 0.15, 0.8, 7, 0.15 and 250, respectively, T is 15, and the average duration v is 86. Then, the std of the random error is 8.6 and 4.3, which are 10 % and 5 % of the average duration, respectively. The average of the absolute values of the error randomly drawn from these sizes then was 6.7 and 3.3, respectively. Figure 5 shows the error-size impacts in terms of the difference between $f(v)$ and $(f(v+6.7) + f(v-6.7))/2$ for error size of 10% and the difference between $f(v)$ and $(f(v+3.3) + f(v-3.3))/2$ for error size of 5%. The error-averaged marginal utility is drawn as bold lines.

The difference between the two is explained in Figure 6. In both figures, the thin, dotted and bold lines respectively denote the true value, the value with negative error, and the value with the positive error.

Figure 5 Average marginal utility curve with exact measurement and with measurement error of 10% (LHS) and 5% (RHS) of average duration

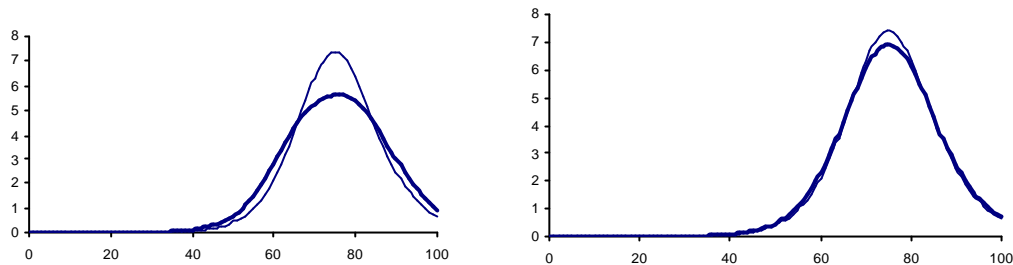
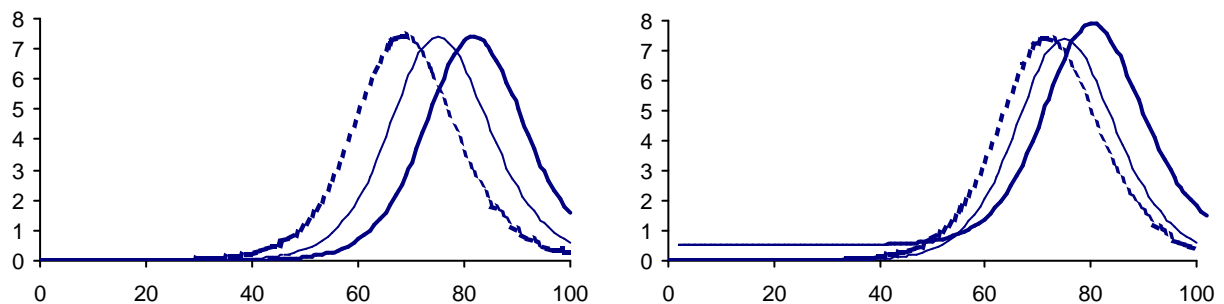


Figure 6 Marginal utility curve with exact measurement and with measurement error of $\pm 10\%$ (LHS) and $\pm 5\%$ (RHS) of average duration



The introduction of a single measurement error changes the α value, and hence, the inflection point of the critical point of the function. As long as the introduced measurement error is from a symmetrical normal distribution, however, the averaged marginal curve maintains the inflection point as before, which also keeps the predicted α values as before. Meanwhile, all other aspects of the functional curve changes as can be seen in Figure 5. As a result, the error introduction from a symmetric (normal) distribution implies the disruption of estimates of \mathbf{b} ,

\mathbf{a}_x and \mathbf{b}_x to some extent, but not α . In fact, throughout the estimation simulation, the α estimate was relatively irrelevant to the measurement error size.

5. Conclusions

This paper has reported some main findings of a study, which aimed at developing an estimation method for deriving the parameters of the Aurora model. Unlike other utility models of time use, Aurora is based on a complex asymmetric S-shaped utility function. While we argue that this specification has some clear theoretical advantages, the estimation of the model becomes highly complex. Before applying a method to real empirical data, we felt it was important to first study the performance of the suggested approach to simulated data.

The suggested estimation uses a combination of searching the solution space, using a tailored genetic algorithm and some theoretical concepts. In particular, a key assumption is that an activity schedule is the result of equalizing the marginal utilities of activities subject to time pressure. The method was specifically developed for the case where duration (time use) data are available.

The results of the simulations suggest that the proposed estimation method performs well on the exact data, and reasonably good on the noisy data but with some exceptions of particular parameters. The simulated noises were time pressure, duration rounding and overall measurement error in duration. Among the simulated errors, the overall measurement error in duration has biggest impacts on the accuracy of the parameter estimation, bigger than the error in measuring the time pressure variables and the systematic rounding in reporting the duration. This suggests the need to attempt to increase the quality of duration data to the extent possible. It also shows the typical problems of non-linear models, especially the interaction between parameter estimates. Considering the latter, we recommend to use a sequential estimation strategy where each parameter is estimated in turn. It did provide the most stable results for the present model, and these results can probably be generalized to similar models.

Having developed this estimation model, we can apply and test the model to real, empirical data. The results of that endeavor will be reported in future publications.

6. References

- Arentze, T.A. and H.J.P. Timmermans (2000) *Albatross: A Learning-Based Transportation Oriented Simulation System*, European Institute of Retailing and Services Studies, The Hague.
- Arentze, T.A., M. Dijst, E. Dugundji, C.H. Joh, L. Kapoen, S. Krijgsman, C. Maat, H.J.P. Timmermans and J. Veldhuisen (2001) *Amadeus: Scope and conceptual development, Paper presented at the 9th World Conference on Transportation Research*, Seoul, July 2001.
- Becker, G. S. (1965) A theory of the allocation of time, *The Economic Journal*, **75** 493-517.
- Davis, L. (ed.) (1991) *Handbook of Genetic Algorithms*, Van Nostrand Reinhold, New York.
- Gärling, T., R. Gillholm and W. Montgomery (1999) The role of anticipated time pressure in activity scheduling, *Transportation*, **26** 173-191.
- Goldberg, D.E. (1989) A gentle introduction to genetic algorithms, in D.E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, 1-25, Addison-Wesley, Amsterdam.
- Herrera F., M. Lozano and J.L. Verdegay (1998) Tackling real-coded genetic algorithms: Operators and tools for behavioral analysis, *Artificial Intelligence Review*, **12** 265-319.
- Joh, C.H., T.A. Arentze and H.J.P. Timmermans (2001a) A theory and simulation model of activity-travel rescheduling behavior, *Paper presented at the 9th World Conference on Transportation Research*, Seoul, July 2001.
- Joh, C.H., T.A. Arentze and H.J.P. Timmermans (2001b) Multidimensional sequence alignment methods for activity-travel pattern analysis: A comparison of dynamic programming and genetic algorithms, *Geographical Analysis*, **33** 247-270.
- Joh, C.H., T.A. Arentze and H.J.P. Timmermans (2002) Modeling individuals' activity-travel rescheduling heuristics: Theory and numerical experiments, *Transportation Research Record*, **1807** 16-25.
- Kitamura, R. (1984) A model of daily time allocation to discretionary out-of-home activities and trips, *Transportation Research B*, **18** 255-266.
- Richards, F.J. (1959) A flexible growth function for empirical use, *Journal of Experimental Botany*, **10** 290-300.

Timmermans, H.J.P., T.A. Arentze and C.H. Joh (2001) Modeling the effects of anticipated time pressure on the execution of activity programs, *Transportation Research Record*, **1752** 8-15.

Wright, A. (1991) Genetic algorithms for real parameter optimization, in G.J.E. Rawlin (ed.), *Foundations of Genetic Algorithms I*, 205-218, Morgan Kaufmann, San Mateo.