

Research Article

Performance Comparison of Mode Choice Optimization Algorithm with Simulated Discrete Choice Modeling

Hyuk-Jae Roh ^{1,2}, Satish Sharma,¹ Prasanta K. Sahu ³ and Babak Mehran ¹

¹Faculty of Engineering and Applied Science, Environmental Systems Engineering, University of Regina, Regina, SK, Canada S4S 0A2

²City of Regina, Regina, SK, Canada S4P 3C8

³Department of Civil Engineering, Birla Institute of Technology and Science Pilani, Rajasthan 333031, India

Correspondence should be addressed to Hyuk-Jae Roh; hyukjae.roh@gmail.com

Received 8 December 2017; Accepted 30 January 2018; Published 9 May 2018

Academic Editor: Luis Carlos Rabelo

Copyright © 2018 Hyuk-Jae Roh et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Until recently, a majority of modeling tasks of transportation planning, especially in discrete choice modeling, is conducted with the help of commercial software and only concerned about the result of parameter estimates to get a policy-sensitive interpretation. This common practice prevents researchers from gaining a systematic knowledge involved in estimation mechanism. In this research, to shed a light on these limited modeling practices, a standard discrete choice model's parameter is estimated using Quasi-Newton method, DFP, and BFGS. Two extended algorithms, called DFP-GSM and BFGS-GSM, are proposed for the first time to overcome the weakness of the Quasi-Newton method. The golden section method (GSM) incorporates a nonlinear programming technique to choose an optimal step size automatically. Partial derivatives of log-likelihood function are derived and coded using Visual Basic Application (VBA). Through extensive numerical evaluation, estimation capability of each proposed estimation algorithms is compared in terms of performance measures. The proposed algorithms show a stable estimation performance and the reasons were studied and discussed. Furthermore, useful insights educated in custom-built modeling are present.

1. Introduction

Discrete choice modeling is widely used across disciplines to predict a certain choice situation through a mathematical inferring of a choice model's parameters [1]. In transportation demand forecasting, this choice analysis method is generally used to describe a variety of choices. Applications of discrete choice theory in various fields are very useful tool for policy analysis and planning. Applications in the transportation sector are very active and diverse and apply to almost every choice situation where alternatives exist, for example, the choice of travel mode [2–7], mode and departure time selection [8, 9], route selection [10–13], and airport selection [14–16]. It is common practice to estimate model parameters by relying on commercially available software.

Until recently, a concern of discrete choice modelers is mainly about to find an interpretable parameter estimates sound in their statistical accuracy level and in their reflected meaning in a policy-sensitive way. Estimation algorithm

and calculation mechanism are actually not concern of researchers. However, as a specification of model becomes more complicated to model a real situation in a persuasive way, and as a more realistic assumption is frequently adopted in a model developing process, researchers (or modelers) might meet difficulties in dealing with modeling tasks. Employing a new probability distribution function (PDF) for describing an unobserved part of a choice model can be prohibited by a limitation of software packages. Choosing and applying a specific algorithm for satisfying a number of occasions in modeling tasks are not easy to practice with a normal software package. In particular, opportunity to have access to all the details of calculation that enable researchers to have a systematic understanding of modeling mechanism is hardly to get unless modeling is in custom-built procedure. A few literature items on this topic can be found in the econometrics field [17–23], but these studies have addressed a different application environment and, therefore, it is difficult to apply these to modeling transportation related decisions

[24, 25]. To tackle this kind of challenges that can occur in model developing, researchers must have a capability of doing modeling with custom-built computer codes that are tailored for each unique modeling case. To be able to reach this level of stage, researchers should be more familiar with a calculation mechanism and a flow to be undertaken before final parameter estimates are taken.

In this research, a standard discrete choice model is estimated using self-made computer code that is developed by using Visual Basic Application (VBA) in EXCEL [26–28]. In order to guarantee convergence of a test model, excellent and robust in convergence properties even for the problems that cannot be solved in satisfactory manner [29, 30], two practically and pedagogically important estimation algorithms are employed to calculate parameter estimates in every iteration stage. These are DFP and BFGS algorithms. More importantly, for the first time, two new estimation algorithms are proposed to improve a way of finding an optimal step size by incorporating a golden section method (GSM) into general routine of both DFP and BFGS. These are DFP-GSM and BFGS-GSM. These four algorithms are compared in terms of performance measures to show a different operational characteristic according to application of several critical factors that are identified and experimented through extensive numerical trials.

The next parts of this document consist of the following. Section 2 is about a model specification assumed within a concept of multinomial logit choice model and data utilized in this research. Section 3 is dedicated to the issues that researchers should have knowledge on in making custom-built modeling procedures. Section 4 deals with the results obtained from extensive numerical experiments that show a relative operational performance among algorithms, responding to changing of criteria. The last section summarizes the results obtained from this research and presents conclusions.

2. Model and Data

As a standard discrete choice model, multinomial logit (MNL) model has been widely applied to describe a choice behavior of decision maker due to its simplicity of its probabilistic choice function, which is the results derived based on two big assumptions, first one is about an utility maximization which explains a choice mechanism based on an utility concept and second one is about an probability distribution function applied to the distribution of unobserved part of total utility. The simple choice function is shown below:

$$P_n(i) = \frac{e^{U_{ni}}}{\sum_j e^{U_{nj}}}, \quad (1)$$

where $P_n(i)$ is a choice probability of mode i and U_{ni} and U_{nj} are a representative utility function of mode i, j respectively. The final form of the multinomial logit model is composed of two parts, a denominator which is the sum of the systematic utility of all the alternatives in the choice set and a numerator which is the systematic utility of the alternative chosen by decision makers.

As a revealed preference (RP) type data, 540 travels interviewed on site for passengers who travel to airport using possible modes are used to estimate a mode choice model. Five alternatives are considered and the representative utility function of each alternative has a specification as follows:

$$U_{\text{car}} = \beta_1 TC_C + \beta_2 TT_C + \beta_3 \text{Acco } m + \beta_4 \text{Sex } m + \beta_5 \text{Age } 340$$

$$U_{\text{taxi}} = \beta_1 TC_T + \beta_2 TT_T + \beta_6 Tasc \quad (2)$$

$$U_{\text{subway}} = \beta_1 TC_S + \beta_2 TT_S + \beta_7 Sasc$$

$$U_{\text{bus}} = \beta_1 TC_B + \beta_2 TT_B + \beta_8 Basc$$

$$U_{\text{limousine}} = \beta_1 TC_L + \beta_2 TT_L + \beta_9 Lasc.$$

3. Components in Custom-Built Discrete Choice Modeling

3.1. Formulating Log-Likelihood Function. To describe estimation process with computer codes using maximum likelihood estimator (MLE), a high-order nonlinear likelihood function containing whole information of the surveyed data is to be built. Likelihood function specific this research is below:

$$L(\beta_1, \beta_2, \dots, \beta_8, \beta_9) = \prod_N \prod_i P_n(i)^{y_{in}} = \prod_N [P_n(C)^{y_{Cn}} \cdot P_n(T)^{y_{Tn}} \cdot P_n(S)^{y_{Sn}} \cdot P_n(B)^{y_{Bn}} \cdot P_n(L)^{y_{Ln}}] \quad (3)$$

By a log transformation, the above equation (3) changed into a tractable form from the mathematical standpoint as shown in (4). By differentiating the equation in first and second order with respect to each parameter, all elements contained in a vector of gradients and the Hessian matrix can be expressed with mathematical expressions that provide a numeric value in iterative estimation process. This estimation process will be discussed in detail in the following section:

$$\begin{aligned} LL(\beta_1, \beta_2, \dots, \beta_8, \beta_9) &= \sum_{n=1}^N \left[y_{Cn} \right. \\ &\cdot \ln \left(\frac{e^{U_{\text{car}}}}{e^{U_{\text{car}}} + e^{U_{\text{taxi}}} + e^{U_{\text{subway}}} + e^{U_{\text{bus}}} + e^{U_{\text{limousine}}}} \right) \\ &+ y_{Tn} \ln \left(\frac{e^{U_{\text{taxi}}}}{e^{U_{\text{car}}} + e^{U_{\text{taxi}}} + e^{U_{\text{subway}}} + e^{U_{\text{bus}}} + e^{U_{\text{limousine}}}} \right) \\ &+ y_{Sn} \ln \left(\frac{e^{U_{\text{subway}}}}{e^{U_{\text{car}}} + e^{U_{\text{taxi}}} + e^{U_{\text{subway}}} + e^{U_{\text{bus}}} + e^{U_{\text{limousine}}}} \right) \\ &+ y_{Bn} \ln \left(\frac{e^{U_{\text{bus}}}}{e^{U_{\text{car}}} + e^{U_{\text{taxi}}} + e^{U_{\text{subway}}} + e^{U_{\text{bus}}} + e^{U_{\text{limousine}}}} \right) \\ &+ y_{Ln} \\ &\cdot \left. \ln \left(\frac{e^{U_{\text{limousine}}}}{e^{U_{\text{car}}} + e^{U_{\text{taxi}}} + e^{U_{\text{subway}}} + e^{U_{\text{bus}}} + e^{U_{\text{limousine}}}} \right) \right]. \quad (4) \end{aligned}$$

TABLE 1: Different way of acquiring the Hessian matrix between algorithms.

Algorithms	NR	BHHH	BHHH-2	SA	DFP	BFGS
Hessian matrix (H_t)	$(-H_{t\text{Real}}^{\text{NR}})^{-1}$	$(H_{t\text{AP}}^{\text{BHHH}})^{-1}$	$(H_{t\text{AP}}^{\text{BHHH-2}})^{-1}$	I^{SA}	$(-H_{t\text{AP}}^{\text{DFP}})$	$(-H_{t\text{AP}}^{\text{BFGS}})$

3.2. Iteration Rule. As the most frequently adopted in commercial statistical package, Quasi-Newton method shows a better performance in running time and is considered to be robust in convergence properties compared to any other algorithms [30]. More importantly, this method might be excellent in convergence properties even for problems that cannot be solved by other algorithms in a satisfactory manner [29]. As noted in the text of Train [30], this unique excellence of the algorithm is the results of selecting a different approach in updating the Hessian matrix, compared to other estimation algorithms such as Newton Raphson, BHHH, BHHH-2, and Steepest Ascent.

A general iteration rule applied in this research for estimating parameters of discrete choice model is shown below:

$$\beta_{t+1} = \beta_t + \lambda (H_t) G_t, \quad (5)$$

where β_t is the parameter estimates after t iterations, β_{t+1} is the parameter estimates after $t + 1$ iterations, λ is a step size (can be assumed by a researcher at the initial time of model parameter estimation), H_t is the Hessian matrix in iteration t in each algorithm, and G_t is a vector of gradient in iteration t . This rule is commonly applied to DFP and BFGS.

As compared in Table 1, each algorithm might be generally tried in custom-built modeling and it is important to know how these algorithms work in estimation process from the pedagogical aspect [30]. In fact, they have been widely adopted as an estimation routine in commercial software packages. The first four algorithms (i.e., Newton Raphson, BHHH, BHHH-2, and Steepest Ascent) are referred from the earlier research of Roh and Khan [25] and Train [30]. For more details refer to the above two researches. In this research, Quasi-Newton method (the last two algorithms of Table 1) is mainly considered in finding parameters of the test model with developed computer codes for this specific purpose. Furthermore, an extension of the method is proposed to present a new technique of searching a step size automatically.

The second iteration rule experimented for the first time in this research shows a difference in a way of adopting a step size (λ_t) which is adjusted repeatedly during estimation of the test model. More details will be discussed in the following section. This iteration rule is commonly applied to DFP-GSM and BFGS-GSM algorithms.

$$\beta_{t+1} = \beta_t + \lambda_t (H_t) G_t. \quad (6)$$

A slight, but notable difference of two iteration rules described in both (5) and (6) is of acquiring a step size (λ , λ_t) during iteration.

3.3. Estimation Algorithms

3.3.1. Davidon-Fletcher-Powell (DFP) Algorithm. The two algorithms, DFP and BFGS, included in the two following sections depend on the description of Greene [29]. Updating rule of approximate Hessian of DFP is introduced below and it is called variable metric algorithm or rank two correction [29], written as

$$H_{t+1(\text{AP})}^{\text{DFP}} = H_t^{\text{DFP}} + \frac{\delta_t \delta_t'}{\delta_t' \gamma_t} - \frac{H_t^{\text{DFP}} \gamma_t \gamma_t' H_t^{\text{DFP}}}{\gamma_t' H_t^{\text{DFP}} \gamma_t}, \quad (7)$$

where $H_{t+1(\text{AP})}^{\text{DFP}}$ is the approximate Hessian matrix of DFP algorithm in $t + 1$ iteration, replacing the real Hessian matrix appeared as $(H_{t\text{Real}}^{\text{NR}})^{-1}$ in Newton Raphson algorithm, H_t^{DFP} is the approximate Hessian matrix in t iteration (if $t = 1$ then $H_t^{\text{DFP}} = I$), in the context of this research I is 9×9 dimension identity matrix, δ_t is equal to the value of $\lambda(-H_t^{\text{DFP}})G_t$ in iteration t , δ_t' is transposed matrix of δ_t , γ_t is equal to the value of $G_{t+1} - G_t$, the difference of value between a gradient vector in iteration $t = 1$, G_{t+1} , and a gradient vector in iteration t , G_t , and γ_t' is transposed matrix of γ_t .

The iteration rule can be written as

$$\beta_{t+1} = \beta_t + \lambda (-H_{t\text{AP}}^{\text{DFP}}) G_t. \quad (8)$$

After sufficient iterations before convergence, $H_{t\text{AP}}^{\text{DFP}}$ is assumed to approximate inverse of real Hessian of NR. In this first DFP method, we can try a variety of λ , a step size, to find the best λ with which the iteration shows the best performance.

3.3.2. Broyden-Fletcher-Goldfarb-Shanno (BFGS) Algorithm. The BFGS algorithm is slightly different in terms of its updating rule of approximate Hessian compared to that of DFP. The updating formula of approximate Hessian is presented below and it shows almost the same formula as for that of DFP except the last term:

$$H_{t+1(\text{AP})}^{\text{BFGS}} = \left(H_t^{\text{BFGS}} + \frac{\delta_t \delta_t'}{\delta_t' \gamma_t} - \frac{H_t^{\text{BFGS}} \gamma_t \gamma_t' H_t^{\text{BFGS}}}{\gamma_t' H_t^{\text{BFGS}} \gamma_t} \right) - v d_t d_t', \quad (9)$$

where $H_{t+1(\text{AP})}^{\text{BFGS}}$ is the approximate Hessian matrix of BFGS algorithm in $t + 1$ iteration, replacing the real Hessian matrix appeared as $(H_{t\text{Real}}^{\text{NR}})^{-1}$ in Newton Raphson algorithm, H_t^{BFGS} is the approximate Hessian matrix in t iteration (if $t = 1$ then $H_t^{\text{BFGS}} = I$), in the context of this research I is 9×9 dimension identity matrix, d_t is equal to the value of $(1/\delta_t' \gamma_t) \delta_t - (1/\gamma_t' H_t^{\text{BFGS}} \gamma_t) H_t^{\text{BFGS}} \gamma_t$, d_t' is transposed matrix of d_t , and v is equal to the value of $\gamma_t' H_t^{\text{BFGS}} \gamma_t$.

The additional term (i.e., $vd_t d_t'$) can be calculated using the information of the first term already provided in parenthesis. As in the case of DFP method, H_{tAP}^{BFGS} would be another approximate Hessian matrix after sufficient iteration. After inputting H_{tAP}^{BFGS} into iteration formula, it has the form as below:

$$\beta_{t+1} = \beta_t + \lambda \left(-H_{tAP}^{BFGS} \right) G_t. \quad (10)$$

In this formula, like the iteration rule of DFP, we can employ several values to find an optimal λ that shows the best estimation performance in terms of several measures such as one iteration time, the number of iterations, and convergence time before convergence.

An strong advantage of the methods, DFP and BFGS, is to accept $\gamma_t = G_{t+1} - G_t$ in updating procedures, which is the difference between continuous gradient in iteration $t + 1$ (i.e., G_{t+1}) and iteration t (i.e., G_t). This means that γ_t might contain more information on the shape of the log-likelihood function and it makes it easier to reach the optimal value that maximizes the log-likelihood function [30].

The following section will suggest the way of finding λ automatically with the help of line search method. In this research, we adopted the golden section method (GSM) as a line search tool. This method was successful in the estimation of parameter estimates.

3.3.3. Golden Section Method (GSM). Especially in custom-built discrete choice modeling, the problem of using estimation algorithm for the estimation of model's parameter is that the researchers have no preliminary knowledge of choosing a step size (λ) that guarantees the best performance of the estimation procedure. To address this problem, we put a new module of line search, which uses golden section method (GSM) between codes of both DFP and BFGS methods. Incorporating the golden section method (GSM) for a line search was tried.

As a line search method which is implemented without using derivatives for the targeted function for finding an optimal value that maximizes or minimizes the function, the basic concept of golden section method (GSM) is to reduce the interval of uncertainty during the search procedure to fix a point for minimizing, in this research, the given function. Let $LL(\beta_{1...9})$ be the function to be minimized and $a \leq \beta \leq b$ is the interval of uncertainty. To find a point for minimizing the function we can exclude portions of the interval of uncertainty that do not contain the optimal point through the line search procedure. After a number of iterations for elimination, infinitesimally small portions are left. By averaging lower and upper bounds of the remaining portions, the optimal point for the given function, the current iteration is calculated.

The following summary of the golden section method (GSM) is of description for minimizing a strictly quasi-convex function over the uncertain interval (a_1, b_1) [31].

(i) *Initialization Step.* Researchers determine an final interval of uncertainty $l > 0$, designating (a_1, b_1) as the initial interval of uncertainty, and let $\gamma_1 = a_1 + (1 - \alpha)(b_1 - a_1)$ and $\mu_1 = a_1 + \alpha(b_1 - a_1)$, where $\alpha = 0.618$ (i.e., reduction ratio; the

TABLE 2: Hessian matrix of DFP-GSM and BFGS-GSM algorithms.

Algorithms	DFP-GSM	BFGS-GSM
Hessian matrix (H_t)	$(-H_{tAP}^{DFP})$	$(-H_{tAP}^{BFGS})$

interval of uncertainty is reduced by a factor of 0.618). Next we calculate $LL(\gamma_1)$ and $LL(\mu_1)$. We let $G = 1$, and go to main routine.

(ii) *Main Routines*

Step 1. If $b_G - a_G < l$, stop. The optimal value is between (a_G, b_G) . Otherwise, if $LL(\gamma_G) > LL(\mu_G)$, go to Step 2; and if $LL(\gamma_G) < LL(\mu_G)$, go to Step 3.

Step 2. Let $a_{G+1} = \mu_G$ and $b_{G+1} = b_G$. Furthermore, let $\gamma_{G+1} = \mu_G$, and let $\mu_{G+1} = a_{G+1} + \alpha(b_{G+1} - a_{G+1})$. Calculate $LL(\mu_{G+1})$ and go to Step 4.

Step 3. Let $a_{G+1} = a_G$ and $b_{G+1} = \gamma_G$. Furthermore, let $\mu_{G+1} = \gamma_G$, and let $\gamma_{G+1} = a_{G+1} + (1 - \alpha)(b_{G+1} - a_{G+1})$. Calculate $LL(\gamma_{G+1})$ and go to Step 4.

Step 4. $G = G + 1$; go to Step 1.

3.3.4. An Application of GSM to This Research. Table 3 shows the example computation of DFP-GSM. In this specific example, we adopt 3 as the interval of uncertainty, and use $b_G - a_G < 0.01$ as stopping criterion. In Table 3, the initial interval of uncertainty is of length $3(l)$, and after 13 iterations involving fourteen observations (see *), the interval of uncertainty is $[0.024402711, 0.033719691]$, so that the minimum point can be calculated to be the midpoint of two values. This midpoint may be considered as an optimal step size λ_t searched by the golden section method in certain iteration before convergence. In particular, in the context of this research, the log-likelihood function of multinomial logit is globally concave [32], and to apply the golden section method (GSM) limited to a strictly quasi-convex function [31], we transform the log-likelihood function made in this research into a symmetry function to x -axis by adding $(-)$ sign, $-LL(\beta_{1...9})$.

3.3.5. DFP-GSM Algorithm. Iteration formula is nearly the same as for the DFP method, other than adopting λ_t instead of λ , which is continuously having different values in every iteration. Line search has an important role of finding optimum value (i.e., λ_t) within the interval of uncertainty. In every iteration, λ_t which minimizes the given log-likelihood function is founded and used in the iteration formula shown below:

$$\beta_{t+1} = \beta_t + \lambda_t \left(-H_{tAP}^{DFP} \right) G_t. \quad (11)$$

3.3.6. BFGS-GSM Algorithm. In the case of BFGS-GSM, all operations are the same as applied to DFP-GSM. The approximate Hessian, H_{tAP}^{BFGS} , can be obtained in the same way as for the BFGS method in Table 2. The iteration rule is as presented below:

$$\beta_{t+1} = \beta_t + \lambda_t \left(-H_{tAP}^{BFGS} \right) G_t. \quad (12)$$

TABLE 3: Example of computation of GSM in the case of DFP-GSM algorithm.

Iteration	a_G	b_G	γ_G	μ_G	LL(γ_G)	LL(μ_G)
1	0	3	1.146	1.854	3493.476457*	5476.740529*
2	0	1.854	0.708228	1.146	2297.447729*	3493.476457
3	0	1.146	0.437772	0.708228	1595.81076*	2297.447729
4	0	0.708228	0.270543096	0.437772	1206.102109*	1595.81076
5	0	0.437772	0.167228904	0.270543096	1005.500045*	1206.102109
6	0	0.270543096	0.103347463	0.167228904	910.8710041*	1005.500045
7	0	0.167228904	0.063881441	0.103347463	871.3625899*	910.8710041
8	0	0.103347463	0.039478731	0.063881441	858.3072898*	871.3625899
9	0	0.063881441	0.024402711	0.039478731	856.576918*	858.3072898
10	0	0.039478731	0.015080875	0.024402711	858.7347619*	856.576918
11	0.015080875	0.039478731	0.024402711	0.03015875	856.576918	856.5491249*
12	0.024402711	0.039478731	0.03015875	0.033719691	856.5491249	856.9727219*
13	0.024402711	0.033719691	0.027961797	0.03015875	856.4520154*	856.5491249

* Values of function evaluated newly in each iteration.

TABLE 4: Serial number assigned to each of four algorithms.

Algorithms	Unique serial number of each algorithm	
Four algorithms	DFP-①	BFGS-②
	DFP-GSM-③	BFGS-GSM-④

All the details of estimation procedures involved in the above iteration rules are coded by using Visual Basic Application (VBA) in EXCEL, Microsoft Inc. All experimental estimations are run with a personal computer (Intel Pentium 4, CPU 2.80 GHz, 512 MB Ram).

4. Experimental Results

Four algorithms mentioned in this research are used to estimate the test model. The same data are input in the estimation in order to compare their estimation performance between algorithms. Three performance measures such as one iteration time, the number of iterations, and convergence time are considered in measuring the performance differences between algorithms.

Various step sizes are selected randomly in the first two algorithms (e.g., DFP and BFGS); all those runs that converged successfully were chosen and summarized. As shown in Table 4, for the last two algorithms (e.g., DFP-GSM and BFGS-GSM), instead of choosing a random step size, we need to guess an uncertain interval. With these conditions, we conducted three categories of experiment as follows. (1) The first category is about choosing different convergence criteria; it is tested in experiments 1.1 and 1.2. (2) The second category is about a subject of different initial guessing of the parameters; this work is tried in experiment 2. (3) The third category is about adopting a new method for guessing of initial Hessian instead of using identity matrix at the initial estimation stage; it is implemented by using a method of both BHHH and BHHH-2.

The four estimation algorithms can be recognized with the unique serial number as shown in Table 4, and three

categories used to compare the estimation performance are presented in Table 5.

4.1. Numerical Experiment 1

4.1.1. An Experiment 1-1 ($CR = 10^{-4}$). As given in Table 5, a vector of the test model's parameters is estimated with a convergence criterion represented by 1 in the table in its mathematical expression, having $CR = 10^{-4}$ as the level of precision of parameter estimates. This convergence criterion is referred from Ben-Akiva and Lermans [1]. The results of the experimental estimation for the four different algorithms are summarized in Table 6 for only experimental trials that converged.

As shown in Table 6 diverse step sizes were tried for the first two algorithms to find the best value of step size for making the log-likelihood function to increase in the fastest way until it reaches the maximum. In the first two algorithms, the step size is fixed at the initial stage of estimation and changed to other values by a researcher.

In the last two algorithms, DFP-GSM, BFGS-GSM, the step size is calculated automatically with the function which is imbedded in the general estimation routine and is operated with the golden section line search method. Instead of choosing step size manually, a researcher must establish interval of uncertainty in which the best step size is searched by the golden section line search method. In this research, eight different uncertain intervals are used and compared. Typically, there are no certain limitations in choosing the uncertain intervals, where choosing the value above 0 is common [31].

Figure 1 shows the estimation performances of the four algorithms in terms of three performance measures. The first two algorithms converge or do not converge in different step sizes. Only converged runs appear in both upper part of Table 6 and left part of Figure 1. In contrast, the last two algorithms converge in the every uncertain interval and their results are presented in lower parts of Table 6 and right part of Figure 1.

TABLE 5: Three categories used in experimental runs.

Experiment scheme	Three categories
Experiment 1	(1) Convergence criteria
Experiment 1-1 (applied to four algorithms)	$\beta_0 = [\beta_{01} = 0, \beta_{02} = 0, \beta_{03} = 0, \dots, \beta_{09} = 0]'$ $\textcircled{1} \left[\frac{1}{k} \sum_{k=1}^k (\beta_{t+1,k} - \beta_{t,k})^2 \right]^{1/2} < CR = 10^{-4}$
Experiment 1-2 (applied to four algorithms)	$\beta_0 = [\beta_{01} = 0, \beta_{02} = 0, \beta_{03} = 0, \dots, \beta_{09} = 0]'$ $\textcircled{2} G'_t(-H_t)^{-1}G_t < CR$ $CR = 10^{-4}, 10^{-5}, 10^{-6}$
Experiment 2 (applied to four algorithms)	(2) Initial guessing of parameters
Experiment 3 (applied to four algorithms)	(3) Initial Hessian matrix
	BHHH, BHHH-2

For the first performance measure, one iteration time, there are somewhat large differences between the first two algorithms and the last two algorithms. This reflects the fact that the last two algorithms take more one iteration time due to the procedure added for conducting a line search. It takes about 1~2 minutes for DFP-GSM and BFGS-GSM to finish one iteration. The values of one iteration time can be used to recognize the operational superiority or inferiority of algorithm because it dominates the overall convergence time for all algorithms, even though it is more reasonable to consider simultaneously both one iteration time and the number of iterations in decision of operational performance of algorithm.

For the second performance measure, the number of iterations, obviously there are big differences both between the step sizes in the same algorithm and between the four different algorithms. In case of DFP, BFGS, for runs converged, as the step size increases, the number of iterations decreases. The result indicates that there is a close relationship between the step size and the number of iterations. Choosing a step size randomly at the initial stage of the estimation can affect the estimation performance negatively if the selection is poor or affect positively if the selection is fortunately good. To address the problem of choosing initial step size and to improve the estimation performance, in this research, a new systematic method, namely, golden section method for a line search, is incorporated in algorithms, DFP and BFGS. As a result, DFP-GSM and BFGS-GSM algorithms show a stable estimation performance in terms of the number of iterations. The iteration numbers presented in the right part of Figure 1 are nearly similar for the all uncertain interval considered in experimental runs except one run of DFP-GSM using $L_{\text{lower}}L_{\text{limit}}(0)$, $U_{\text{pper}}L_{\text{limit}}(1)$ as an uncertain interval. Stated simply, the DFP-GSM and BFGS-GSM algorithms compared to the DFP and BFGS in Figure 1 definitely make a decrease in the number of iterations. Moreover, the iteration number converges to around a certain value: these new algorithms are stable in terms of the number of iterations.

The third performance measure, the convergence time, is used to compare the estimation performance of the four algorithms. This measure is the most important measure in order to decide the excellence of an algorithm because the operational time before convergence is of much interest to researchers; especially in large scale modeling it can be very critical.

As a Quasi-Newton method, DFP and BFGS algorithms use a similar updating process for the Hessian matrix as explained in earlier section. The BFGS algorithm takes less convergence time (e.g., $\lambda = 1/4$ or $\lambda = 1/16$) or more convergence time (e.g., $\lambda = 1/8$, $\lambda = 1/2$) than DFP when compared in the same step size. Key point is that, even though the number of iterations is much less in BFGS than that in DFP for the estimation using the same step sizes, the difference in convergence time between two algorithms is not significant. This phenomenon might be attributed to the updating process of the Hessian matrix: BFGS uses a more complicated process than that of DFP and it takes more time in finishing one iteration. The last two algorithms used for the first time in this research, namely, DFP-GSM and BFGS-GSM, show stable estimation performance. In other words, the convergence times are distributed more evenly than other two algorithms. The BFGS-GSM shows slightly better performance than the DFP-GSM in terms of convergence time.

In conclusion, in experiment 1, compared to DFP and BFGS, DFP-GSM and BFGS-GSM show stable performance in terms of both the number of iterations and the convergence time irrespective of the choice of any uncertain interval. DFP and BFGS show somewhat large differences in estimation performance, depending on the selection of different step sizes at the initial stage of estimation.

As shown in Figure 2, the line graphs of the experimental estimations are drawn to show the trace of log-likelihood function values during the iterations and moreover to compare clearly the different convergence behaviors both between step sizes and uncertain interval in each algorithm. Table 7

TABLE 6: The results of four estimation algorithms (exp. 1-1) (CR = 10^{-4}).

Algorithms & performance measures	Step size (λ)							
	1/32	1/16	1/8	1/4	1/2	1	2	4
DFP-(1)								
One iteration time	-	0:00:23	0:00:15	0:00:15	0:00:14	-	-	-
Number of iterations	-	224	193	155	115	-	-	-
Convergence time	-	1:25:04	0:47:26	0:38:49	0:27:32	-	-	-
BFGS-(2)								
One iteration time	0:00:16	0:00:28	0:00:39	0:00:21	0:00:45	0:00:13	-	-
Number of iterations	268	162	107	77	52	42	-	-
Convergence time	1:33:28	1:15:04	1:08:50	0:26:15	0:38:52	0:09:16	-	-
Algorithms & performance measures								
	LL(0), UL(1)	LL(0), UL(2)	LL(0), UL(3)	LL(0), UL(4)	LL(0), UL(5)	LL(0), UL(6)	LL(0), UL(7)	LL(0), UL(8)
Interval of uncertainty								
DFP-GSM-(3)								
One iteration time	0:00:44	0:01:09	0:00:59	0:01:04	0:01:44	0:01:52	0:01:54	0:01:55
Number of iterations	138	23	27	26	24	24	24	23
Convergence time	1:40:55	0:26:29	0:26:29	0:27:40	0:41:38	0:44:46	0:45:37	0:44:01
BFGS-GSM-(4)								
One iteration time	0:00:48	0:00:48	0:00:53	0:00:56	0:01:02	0:01:06	0:01:06	0:02:05
Number of iterations	38	25	21	20	20	19	19	19
Convergence time	0:30:07	0:19:45	0:18:33	0:18:42	0:20:30	0:20:58	0:20:55	0:39:35

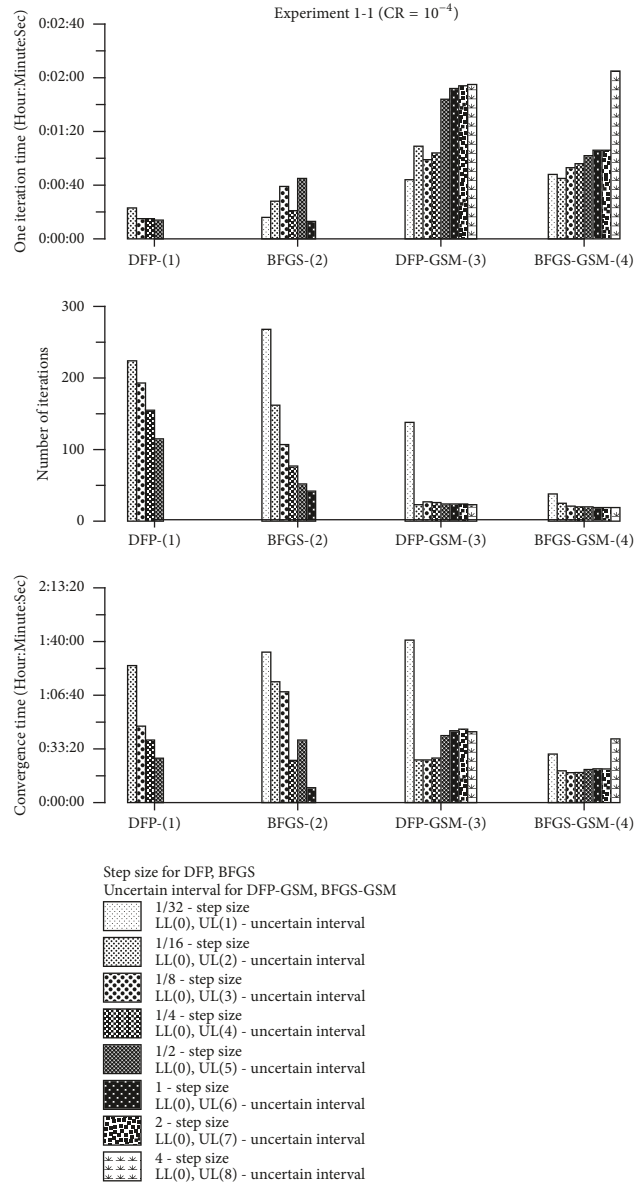


FIGURE 1: The results of estimation for four algorithms (exp. 1-1) (CR = 10^{-4}).

summarizes critical values in iterations such as starting log-likelihood value and convergence log-likelihood value and number of iterations according to both different step sizes and uncertain interval. As shown in Table 7, all starting log-likelihood values are exactly the same for the all four algorithms and both for any step sizes and any uncertain intervals within each algorithm because for exp. 1-1 (CR = 10^{-4}) the initial guess of parameters is assumed to be zero, $\beta_k = [\beta_1 = 0, \beta_2 = 0, \beta_3 = 0, \dots, \beta_9 = 0]^T$.

On the contrary, convergence values are slightly different both between algorithms and between step sizes for the first two algorithms. For the last two algorithms, they are exactly the same in their values for almost all uncertain intervals except some of them.

In the case of the DFP algorithm, there is a large difference in its convergence behavior compared to the four estimation

algorithms discussed in the research of Roh and Khan [25]. They used exactly the same experimental setting except using different four estimation algorithms (i.e., Newton Raphson, BHHH, BHHH-2, and Steepest Ascent). The most obvious difference is that the line graphs for all step sizes, except 1/16, show much bigger fluctuations than the four algorithms [25] in the initial part of iterations without a steady increase. This phenomenon might be assumed as a result of the mechanism used for updating the approximate Hessian matrix of DFP algorithm. Train [30] indicates that this kind of algorithm, including BFGS algorithm, uses information obtained at more than one point on the log-likelihood function for a calculation of approximate Hessian matrix in every iteration. The step size of 1/2 shows the best performance with 115 iterations and its critical values are presented in Table 7.

Experiment 1-1 (CR = 10^{-4})

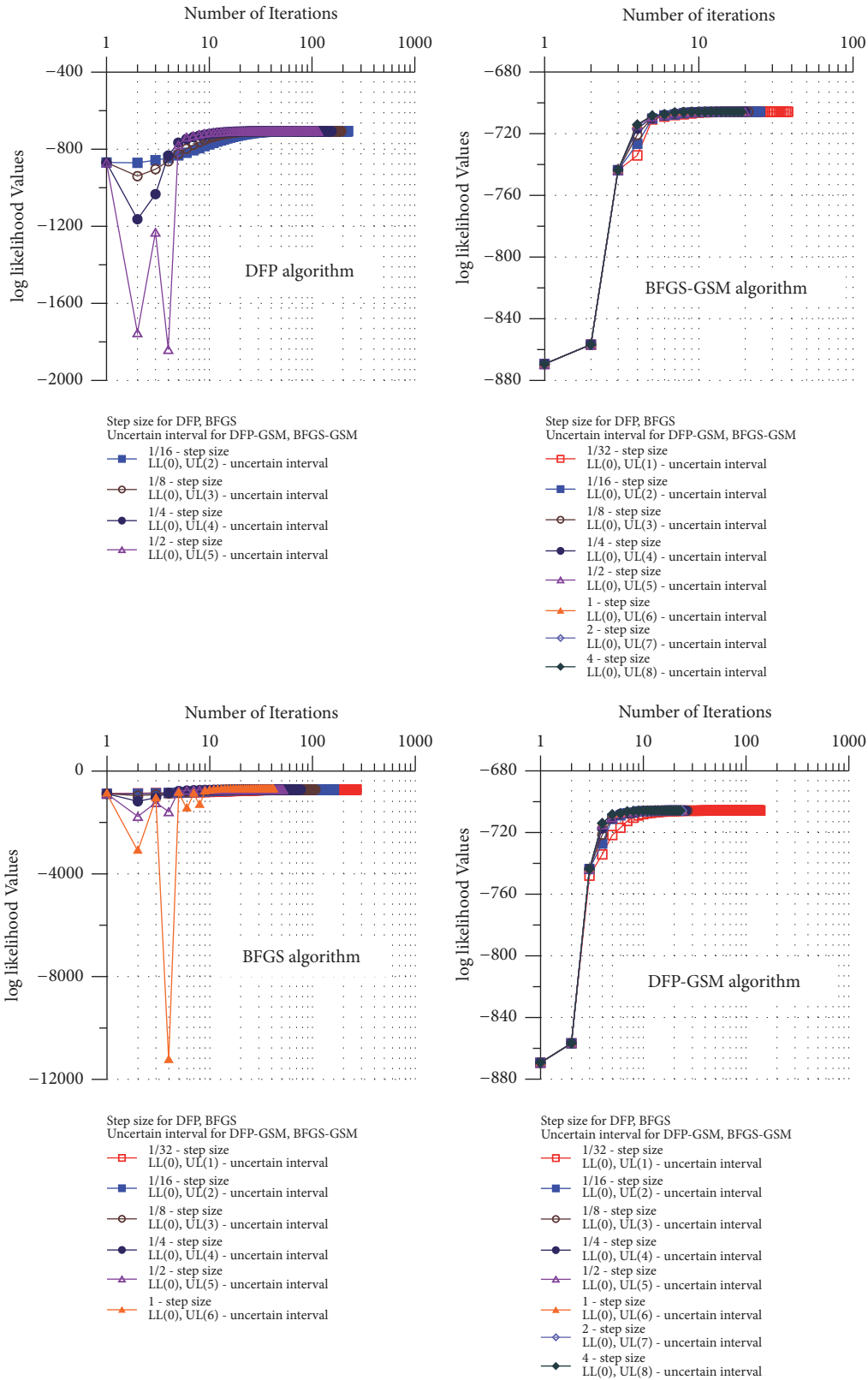


FIGURE 2: Variances of log-likelihood values for four algorithms (exp. 1-1) (CR = 10^{-4}).

TABLE 7: Log-likelihood values obtained from four algorithm (exp. 1-1) (CR = 10^{-4}).

Algorithms & log-likelihood values	Step size (λ)							
	1/32	1/16	1/8	1/4	1/2	1		
DFP-(1)								
Starting	-	-869.0964727	-869.0964727	-869.0964727	-869.0964727	-		
Convergence	-	-705.5398009	-705.5394156	-705.5396982	-705.5406805	-		
Iteration num.	-	224	193	155	115	-		
BFGS-(2)								
Starting	-869.0964727	-869.0964727	-869.0964727	-869.0964727	-869.0964727	-869.0964727		
Convergence	-705.5387064	-705.5384933	-705.5384097	-705.5383265	-705.5383351	-705.5383081		
Iteration num.	268	162	107	77	52	42		
Interval of uncertainty								
Algorithms & log-likelihood values	LL(0), UL(1)	LL(0), UL(2)	LL(0), UL(3)	LL(0), UL(4)	LL(0), UL(5)	LL(0), UL(6)	LL(0), UL(7)	LL(0), UL(8)
DFP-GSM-(3)								
Starting	-869.0964727	-869.0964727	-869.0964727	-869.0964727	-869.0964727	-869.0964727	-869.0964727	-869.0964727
Convergence	-705.538356	-705.5384214	-705.5383015	-705.5383015	-705.5383026	-705.5383015	-705.5383015	-705.5383015
Iteration num.	138	23	27	26	24	24	24	23
BFGS-GSM-(4)								
Starting	-869.0964727	-869.0964727	-869.0964727	-869.0964727	-869.0964727	-869.0964727	-869.0964727	-869.0964727
Convergence	-705.5383016	-705.5383015	-705.5383015	-705.5383016	-705.5383015	-705.5383015	-705.5383015	-705.5383015
Iteration num.	38	25	21	20	20	19	19	19

In the case of the BFGS algorithm, like the DFP algorithm, there are large fluctuations for all step sizes except 1/32 and 1/16. Moreover, a similar reason used in DFP can be applied to explain this phenomenon. Overall, there is only a slight difference in the method to calculate updated approximate Hessian matrix for two algorithms, DFP and BFGS. One more term is added to BFGS algorithm as was shown in earlier section. A run with step size 1 outperforms the other runs, as can be seen in Table 6. It can be noted that it experiences an extreme change in value of log-likelihood at the initial iterations.

In the last two algorithms, the use of a variety of step sizes is not needed because the golden section line search method (GSM) takes over the role of choosing an optimum step size using an automatic method. Another task arises instead, making an interval of uncertainty. In this research, eight different uncertain intervals are considered. It should be noted that there are no rules and no guidance is available from literature regarding the choice of uncertain interval. Temporarily, eight different intervals characterized by an increasing of the same amount, namely 1, were applied to simply examine the convergence behavior that might be caused with different uncertain intervals. Figure 2 shows the convergence behavior of the DFP-GSM and BFGS-GSM algorithms according to various intervals. All line graphs show almost same convergence profile and thus they are drawn in the same line irrespective of uncertain intervals used temporarily. Almost the same results were obtained from DFP-GSM and BFGS-GSM algorithm, as presented in Figure 2.

4.1.2. An Experiment 1-2 (CR = 10⁻⁴). In an experiment 1-2, the test model estimates a vector of parameters based on different convergence criterion in both its mathematical form and the degree of stopping criterion. The mathematical form of convergence criterion used for this experiment is referred to from Besley [17]. The different degree of stopping criterion applies gradually from CR = 10⁻⁴ to CR = 10⁻⁶ in order to examine the impact of applying different degree of convergence precision on the operational behavior in discrete choice model estimation. Table 8 shows the results of the experimental estimation for the four different algorithms for all experimental trials that converged. Figure 3 shows the results graphically in terms of three performance measures.

All conditions and environments applied in the experimental estimation are exactly the same as experiment 1-1 except the formulation of the convergence criterion written in the form of $G_t'(-H_t)^{-1}G_t < 10^{-4}$.

For all algorithms, the estimation performance is improved as compared to experiment 1-1 in terms of three performance measures. This phenomenon can be seen in Table 8. There are significant improvements for the first two algorithms for all step sizes and for all performance measures considered. For the last two algorithms, all uncertain intervals except both $L_{\text{lower}}L_{\text{limit}}(0)$, $U_{\text{pper}}L_{\text{limit}}(2)$ and $L_{\text{lower}}L_{\text{limit}}(0)$, $U_{\text{pper}}L_{\text{limit}}(3)$ show an improvement of performance in terms of convergence time. Another important point is that the last two algorithms show a better performance in terms of both the number of iterations and

the convergence time as the uncertain intervals increase gradually by an amount of unit, 1.

In conclusion, when a new convergence criterion that has a different mathematical form is applied, all algorithms show an improvement in terms of three performance measures compared to the case of experiment 1-1.

As shown in Table 9 for the first two algorithms and for any step sizes used within each algorithm, all calculated starting log-likelihood values are exactly the same due to $\beta_k = [\beta_1 = 0, \beta_2 = 0, \beta_3 = 0, \dots, \beta_9 = 0]'$. The same condition imposed as an initial guess of starting points in experiments 1-1, 1-2, and 3. On the contrary, convergence values are slightly different among the step sizes used in each algorithm and the differences are so small that it can be considered as being equal. As shown in Table 9 for the last two algorithms, all starting log-likelihood values are exactly the same due to the same reason explained for the first two algorithms. In case of the convergence log-likelihood values, slight differences exist among uncertain intervals but these are also so small that these can be considered to be converged at the same value even though it is not exactly the same.

As we can see by comparing two figures, Figure 2 and Figure 4, the most important point we can find is that the last part of iteration procedures has a different profile in graph. That is, the overlap parts observed in experiment 1-1 are less and simplified in experiment 1-2. It means that a researcher can control the termination timing by changing the convergence criterion within a scientific way, of course, if or only if the new given convergence criterion guarantees the degree of accuracy of parameter estimates up to the level which a researcher expects it must be.

4.1.3. An Experiment 1-2 (CR = 10⁻⁵). In this experiment, the applied convergence criterion is $G_t'(-H_t)^{-1}G_t < 10^{-5}$. The degree of precision of parameter estimates is changed to more stringent value, from $G_t'(-H_t)^{-1}G_t < 10^{-4}$ to $G_t'(-H_t)^{-1}G_t < 10^{-5}$. Information on this criterion was obtained from an econometric modeling study of Besley [17] and Belsley [18].

The results of the experimental estimation for the algorithms are summarized in Table 10 in the same pattern and style as used in previous presentations. Detailed results are compared graphically in Figure 5 in terms of three performance measures.

Generally, for all four algorithms, the estimation performance of the current experiment 1-2 (CR = 10⁻⁵) has improved compared to that of experiment exp. 1-1 (CR = 10⁻⁴) but has slightly deteriorated as compared to that of experiment 1-2 (CR = 10⁻⁴). The poor performance observed in this experiment can be attributed to the stopping criterion moved in the stringent direction. It must be the fact that a more stringent stopping criterion needs more calculation time to find a vector of parameter estimates satisfying the stringent stopping criterion.

In conclusion, when a new more strict convergence criterion is applied as a new stopping standard, all algorithms show better estimation performance than experiment 1-1 (CR = 10⁻⁴) but present poorer performance than experiment 1-2 (CR = 10⁻⁴). If a more stringent stopping criterion is applied, the time to convergence tends to increase.

TABLE 8: The results of four estimation algorithms (exp. 1-2) ($CR = 10^{-4}$).

Algorithms & performance measures	Step size (λ)							
	1/32	1/16	1/8	1/4	1/2	1	2	4
DFP-(1)								
One iteration time	-	0:00:17	0:00:16	0:00:23	0:00:22	-	-	-
Number of iterations	-	79	56	42	29	-	-	-
Convergence time	-	0:22:03	0:14:35	0:15:56	0:10:30	-	-	-
BFGS-(2)								
One iteration time	0:00:21	0:00:16	0:00:22	0:00:21	0:00:21	0:00:21	-	-
Number of iterations	150	78	49	32	22	20	-	-
Convergence time	0:52:58	0:20:54	0:17:25	0:10:56	0:07:23	0:06:50	-	-
Algorithms & Performance measures								
	LL(0), UL(1)	LL(0), UL(2)	LL(0), UL(3)	LL(0), UL(4)	LL(0), UL(5)	LL(0), UL(6)	LL(0), UL(7)	LL(0), UL(8)
DFP-GSM-(3)								
One iteration time	0:03:02	0:03:11	0:03:36	0:01:27	0:01:07	0:01:34	0:01:34	0:01:34
Number of iterations	23	11	11	10	10	8	8	8
Convergence time	1:09:47	0:34:55	0:39:24	0:14:26	0:11:03	0:12:27	0:12:27	0:12:26
BFGS-GSM-(4)								
One iteration time	0:01:05	0:01:13	0:01:13	0:01:09	0:01:09	0:01:03	0:01:13	0:01:12
Number of iterations	15	11	11	10	10	10	9	9
Convergence time	0:16:17	0:13:17	0:13:21	0:11:26	0:11:25	0:12:07	0:10:48	0:10:45

TABLE 9: Log-likelihood values obtained from four algorithm (exp. 1-2) (CR = 10^{-4}).

Algorithms & log-likelihood values	Step size (λ)							
	1/32	1/16	1/8	1/4	1/2	1	1	1
DFP-(1)								
Starting	-	-869.0964727	-869.0964727	-869.0964727	-869.0964727	-	-	-
Convergence	-	-705.5861209	-705.6160738	-705.6565252	-705.6940834	-	-	-
Iterations num.	-	79	56	42	29	-	-	-
BFGS-(2)								
Starting	-869.0964727	-869.0964727	-869.0964727	-869.0964727	-869.0964727	-869.0964727	-869.0964727	-869.0964727
Convergence	-705.5772393	-705.5855181	-705.5888418	-705.5979521	-705.6348579	-705.6311626	-705.6311626	-705.6311626
Iteration num.	150	78	49	32	22	20	20	20
Algorithms & log-likelihood values								
Interval of uncertainty								
	LL(0), UL(1)	LL(0), UL(2)	LL(0), UL(3)	LL(0), UL(4)	LL(0), UL(5)	LL(0), UL(6)	LL(0), UL(7)	LL(0), UL(8)
DFP-GSM-(3)								
Starting	-869.0964727	-869.0964727	-869.0964727	-869.0964727	-869.0964727	-869.0964727	-869.0964727	-869.0964727
Convergence	-705.7026844	-705.7337937	-705.7334852	-705.775286	-705.7586887	-705.9187437	-705.8917489	-705.9084545
Iteration num.	23	11	11	10	10	8	8	8
BFGS-GSM-(4)								
Starting	-869.0964727	-869.0964727	-869.0964727	-869.0964727	-869.0964727	-869.0964727	-869.0964727	-869.0964727
Convergence	-705.5879102	-705.6275413	-705.5961032	-705.6161016	-705.6100727	-705.6006594	-705.6241932	-705.6279414
Iteration num.	15	11	11	10	10	10	9	9

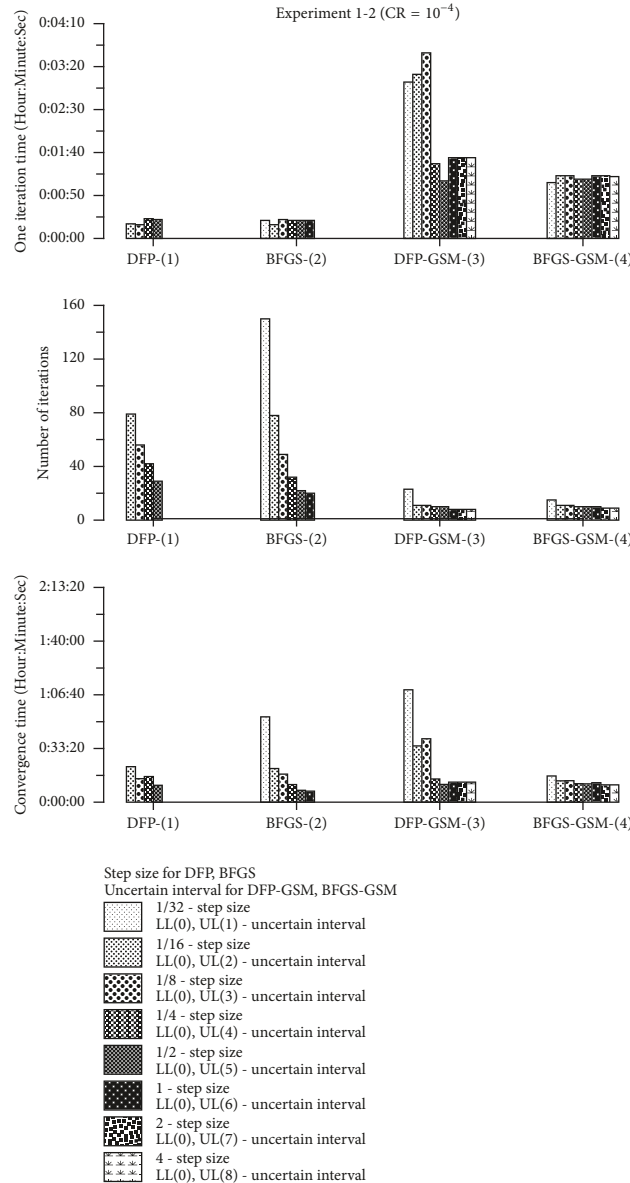


FIGURE 3: The results of estimation for four algorithms (exp. 1-2) (CR = 10⁻⁴).

Moreover, the mathematical formula of stopping criterion used in experiment 1-1 takes more convergence time than that of experiment 1-2.

For all estimation results, the graphs are drawn in order to show a variation of the log-likelihood function values following the iterations and moreover to compare clearly the different convergence behaviors between step sizes used in each algorithm. Table 11 summarizes critical values such as starting log-likelihood value and convergence log-likelihood value and the number of iterations due to the different step sizes. As shown in Table 11, all calculated starting log-likelihood values are exactly the same for the all four algorithms and for any step sizes used within each algorithm due to $\beta_k = [\beta_1 = 0, \beta_2 = 0, \beta_3 = 0, \dots, \beta_9 = 0]^T$, which implies the same condition imposed as an initial guess of starting

points in experiments 1-1, experiment 1-2, and experiment 3. On the other hand, convergence values are slightly different among the step sizes only below the decimal point and so the difference can be ignored and all runs are considered to be converged at the same log-likelihood value.

The convergence behavior is identical with experiments 1-1 (CR = 10⁻⁴) and 1-2 (CR = 10⁻⁴) except the fact that slight differences happen over the three performance measures. These experimental results are clearly presented in Figure 6. Overall, the phenomena observed are that, with decreasing step size, the slope also becomes small but the number of iterations rises.

We can clearly identify the best performance among the runs performed using various step sizes for the first two algorithms but in the last two algorithms almost all graph

Experiment 1-2 (CR = 10^{-4})

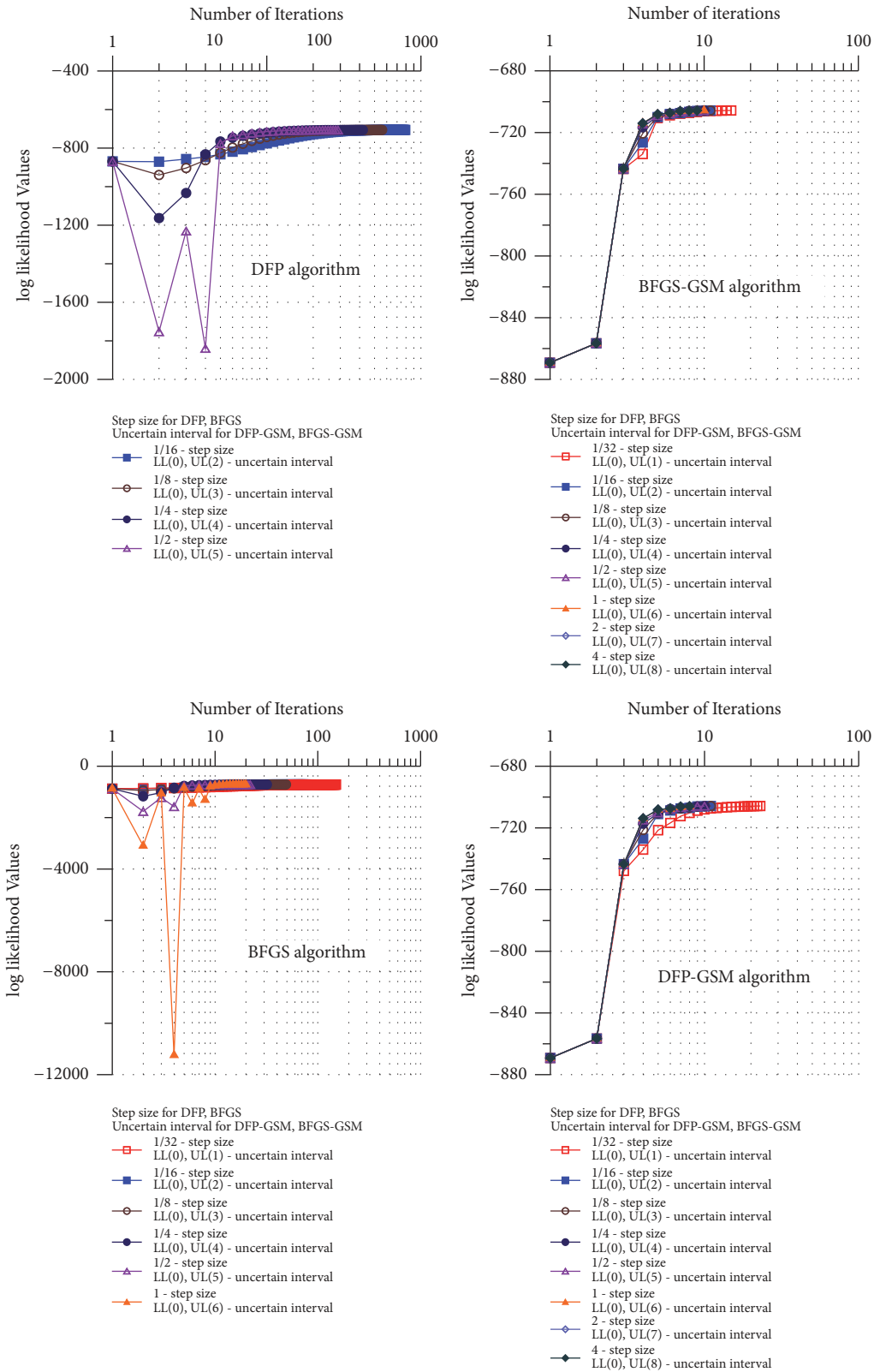


FIGURE 4: Variances of log-likelihood values for four algorithms (exp. 1-2) (CR = 10^{-4}).

TABLE 10: The results of four estimation algorithms (exp. 1-2) (CR = 10^{-5}).

Algorithms & performance measures	Step size (λ)							
	1/32	1/16	1/8	1/4	1/2	1	2	4
DFP-(1)								
One iteration time	-	0:00:20	0:00:19	0:00:18	0:00:20	-	-	-
Number of iterations	-	110	83	62	45	-	-	-
Convergence time	-	0:36:37	0:26:21	0:19:03	0:14:32	-	-	-
BFGS-(2)								
One iteration time	0:00:19	0:00:20	0:00:20	0:00:19	0:00:19	0:00:19	-	-
Number of iterations	212	126	64	42	29	25	-	-
Convergence time	0:68:41	0:41:25	0:21:00	0:13:31	0:09:13	0:07:48	-	-
Algorithms & performance measures								
	LL(0), UL(1)	LL(0), UL(2)	LL(0), UL(3)	LL(0), UL(4)	LL(0), UL(5)	LL(0), UL(6)	LL(0), UL(7)	LL(0), UL(8)
DFP-GSM (3)								
One iteration time	0:03:02	0:03:12	0:03:16	0:01:16	0:03:23	0:01:53	0:01:53	0:01:50
Number of iterations	33	15	15	14	14	13	13	13
Convergence time	1:40:14	0:47:57	0:48:48	0:17:44	0:47:15	0:24:26	0:24:27	0:23:50
BFGS-GSM-(4)								
One iteration time	0:01:21	0:01:26	0:01:34	0:01:30	0:01:33	0:01:33	0:01:25	0:01:32
Number of iterations	19	14	13	13	12	12	11	11
Convergence time	0:25:31	0:19:55	0:20:16	0:19:18	0:18:36	0:18:27	0:15:21	0:16:47

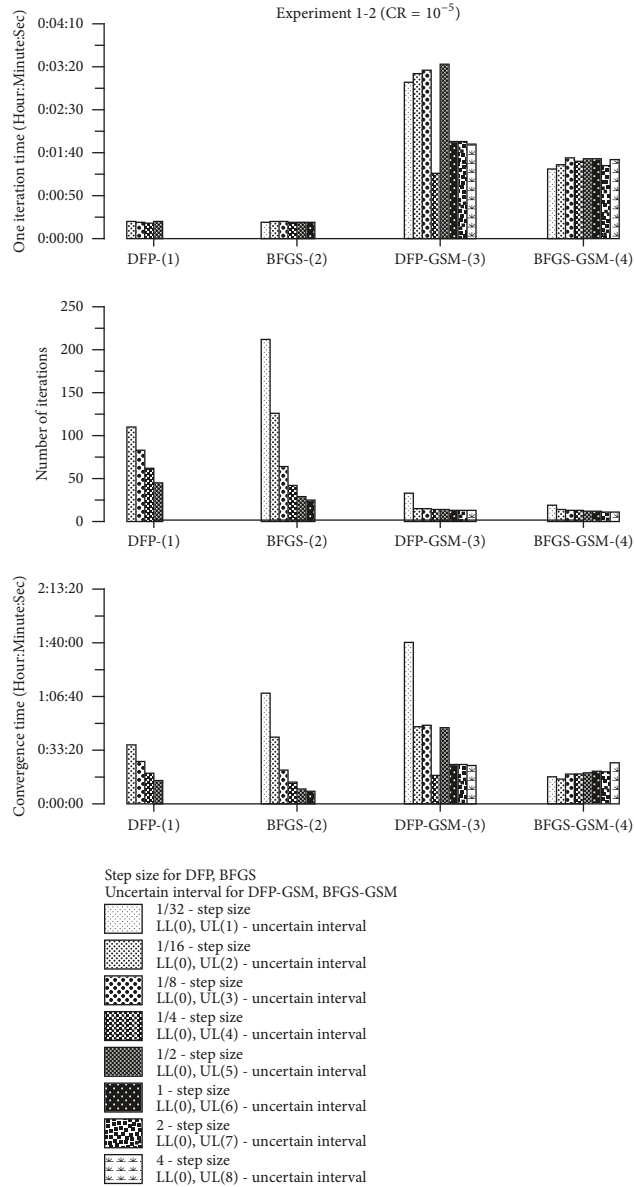


FIGURE 5: The results of estimation for four algorithms (exp. 1-2) (CR = 10⁻⁵).

drawn for each uncertain interval look to be drawn in the same line graph (see Figure 6). Also, the profile of graph at the last iteration is not as long as that of experiment 1-1 (CR = 10⁻⁴) but as compared with experiment 1-2 (CR = 10⁻⁴), the overlapped part becomes longer. It means that more iteration is conducted in each algorithm to satisfy a new stringent criterion charged for stopping the iteration process. It also means that the iteration number and convergence time can be adjusted to become suitable for each specific modeling work or for maintaining the accuracy of parameters within the desired accuracy demand.

4.1.4. An Experiment 1-2 (CR = 10⁻⁶). Here, the applied convergence criterion is $G_t'(-H_t)^{-1}G_t < 10^{-6}$. The degree of precision of parameter estimates is changed to more stringent

value from $G_t'(-H_t)^{-1}G_t < 10^{-5}$ to $G_t'(-H_t)^{-1}G_t < 10^{-6}$. Besley [17] is credited for showing the benefits of using this criterion in econometric modeling.

The results obtained from the experimental estimation for the four algorithms are summarized in Table 12, and the detailed results are presented graphically in Figure 7 in terms of three performance measures.

As presented in Table 12, generally, for all four algorithms, the estimation performance of the current experiment 1-2 (CR = 10⁻⁶) has deteriorated as compared to that of experiments 1-2 (CR = 10⁻⁴) and experiment 1-2 (CR = 10⁻⁵) and even the performance is below that of experiment 1-1 (CR = 10⁻⁴). The reason for the worst performance witnessed in this experiment 1-2 (CR = 10⁻⁶), as was the situation observed in experiment 1-2 (CR = 10⁻⁵), can be

TABLE II: Log-likelihood values obtained from four algorithm (exp. 1-2) ($CR = 10^{-5}$).

Algorithms & log-likelihood values	Step size (λ)							
	1/32	1/16	1/8	1/4	1/2	1	1	1
DFP-(1)								
Starting	-	-869.0964727	-869.0964727	-869.0964727	-869.0964727	-	-	-
Convergence	-	-705.5515777	-705.5521877	-705.5563912	-705.5649595	-	-	-
Iteration num.	-	110	83	62	45	-	-	-
BFGS-(2)								
Starting	-869.0964727	-869.0964727	-869.0964727	-869.0964727	-869.0964727	-869.0964727	-869.0964727	-869.0964727
Convergence	-705.5433026	-705.5435922	-705.5504568	-705.5505559	-705.5514638	-705.5638547	-705.5638547	-705.5638547
Iteration num.	212	126	64	42	29	25	25	25
Algorithms & log-likelihood values								
	LL(0), UL(1)	LL(0), UL(2)	LL(0), UL(3)	LL(0), UL(4)	LL(0), UL(5)	LL(0), UL(6)	LL(0), UL(7)	LL(0), UL(8)
DFP-GSM-(3)								
Starting	-869.0964727	-869.0964727	-869.0964727	-869.0964727	-869.0964727	-869.0964727	-869.0964727	-869.0964727
Convergence	-705.5585592	-705.5623115	-705.5614377	-705.5583746	-705.5494868	-705.551434	-705.5497142	-705.5494614
Iteration num.	33	15	15	14	14	13	13	13
BFGS-GSM-(4)								
Starting	-869.0964727	-869.0964727	-869.0964727	-869.0964727	-869.0964727	-869.0964727	-869.0964727	-869.0964727
Convergence	-705.5495373	-705.5515069	-705.5498357	-705.5494556	-705.5509704	-705.5498048	-705.5505166	-705.550681
Iteration num.	19	14	13	13	12	12	11	11

Experiment 1-2 (CR = 10⁻⁵)

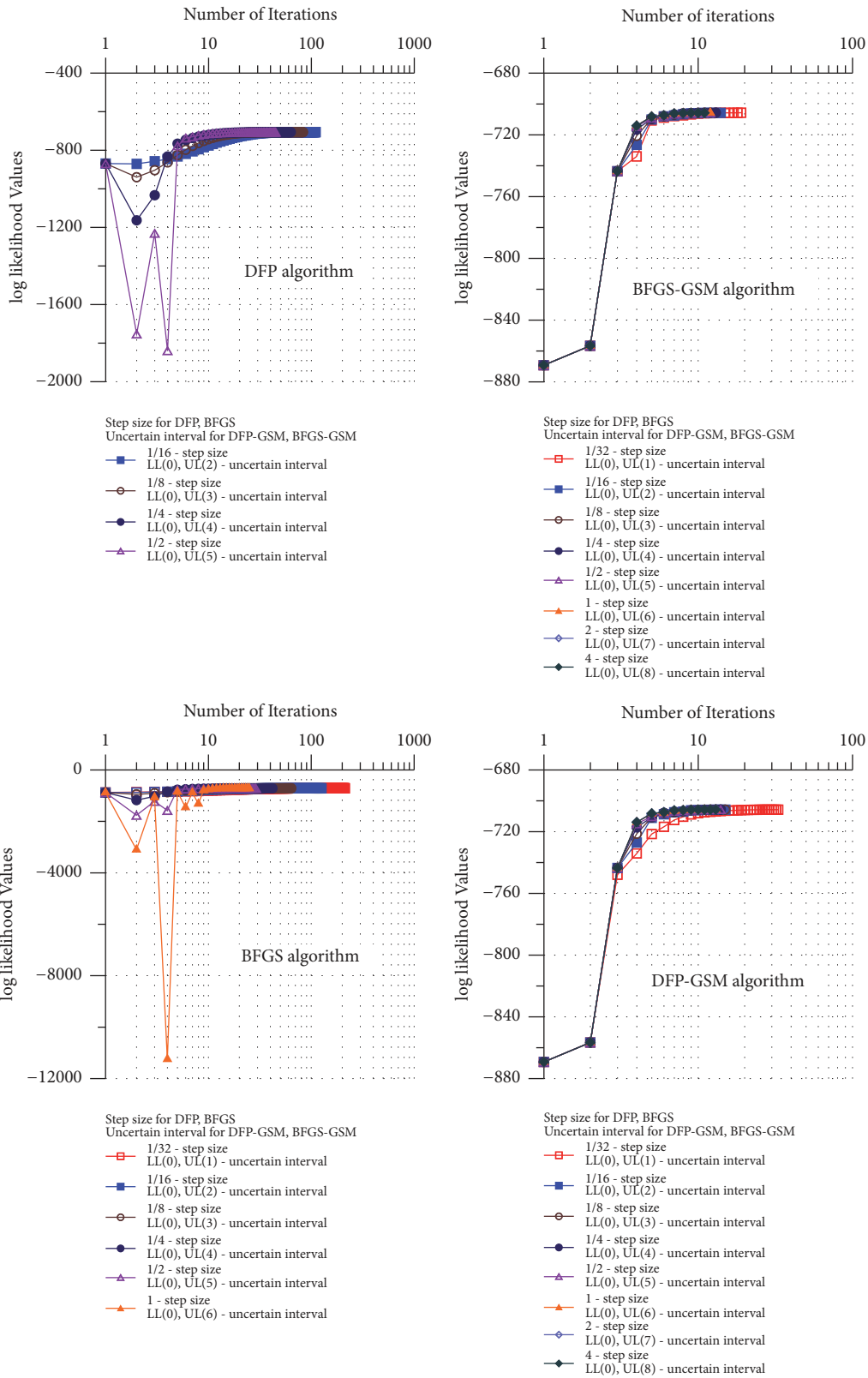


FIGURE 6: Variances of log-likelihood values for four algorithms (exp. 1-2) (CR = 10⁻⁵).

TABLE 12: The results of four estimation algorithms (exp. 1-2) (CR = 10^{-6}).

Algorithms & performance measures	Step size (λ)							
	1/32	1/16	1/8	1/4	1/2	1	2	4
DFP-(1)								
One iteration time	-	0:00:20	0:00:19	0:00:20	0:00:26	-	-	-
Number of iterations	-	216	139	95	69	-	-	-
Convergence time	-	1:21:11	0:43:39	0:31:45	0:29:20	-	-	-
BFGS-(2)								
One iteration time	0:00:17	0:00:19	0:00:18	0:00:19	0:00:19	0:00:18	-	-
Number of iterations	274	158	100	66	48	36	-	-
Convergence time	1:16:01	0:49:15	0:30:39	0:20:22	0:14:55	0:10:44	-	-
Algorithms & performance measures								
	LL(0), UL(1)	LL(0), UL(2)	LL(0), UL(3)	LL(0), UL(4)	LL(0), UL(5)	LL(0), UL(6)	LL(0), UL(7)	LL(0), UL(8)
Interval of uncertainty								
DFP-GSM-(3)								
One iteration time	0:02:27	0:02:51	0:01:30	0:01:12	0:01:34	0:01:40	0:01:38	0:01:49
Number of iterations	42	17	17	16	14	14	14	13
Convergence time	1:43:06	0:48:24	0:25:26	0:19:08	0:21:53	0:20:10	0:22:47	0:23:26
BFGS-GSM-(4)								
One iteration time	0:01:16	0:01:21	0:01:28	0:01:22	0:01:22	0:01:28	0:01:29	0:01:29
Number of iterations	26	15	16	13	13	12	14	14
Convergence time	0:32:51	0:20:13	0:23:29	0:17:40	0:17:32	0:17:26	0:20:45	0:20:32

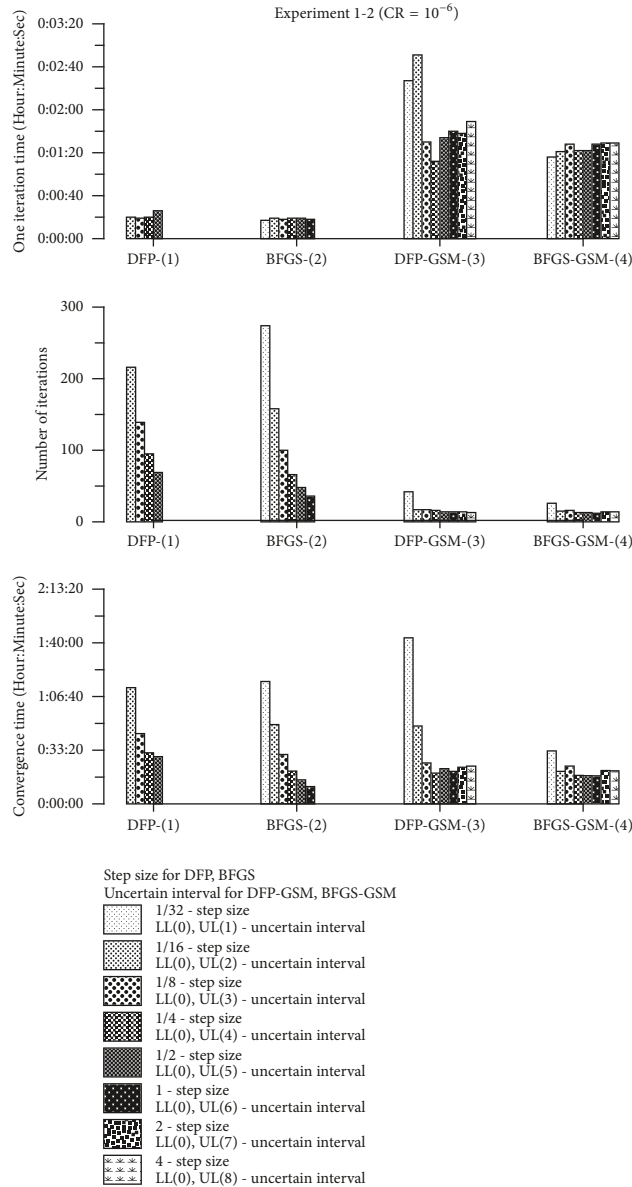


FIGURE 7: The results of estimation for four algorithms (exp. 1-2) (CR = 10⁻⁶).

attributed to the most stringent stopping criterion applied in this experiment. The reason is the fact that the most stringent stopping criterion needs much more calculation time than for runs having a less stringent stopping criterion. Finding a vector of parameter estimates that satisfies this stringent stopping criterion is a time consuming process and it may not be worthwhile. Researchers can take moderate procedures and controlling measure according to their expectation and specific purpose in doing their unique modeling.

In conclusion, when the most strict convergence criterion, specifically in this research, is applied as a new stopping measure, all algorithms show a worse estimation performance than all experiments tried previously. Applying a more stringent stopping criterion causes an increase in convergence time in discrete choice model estimation.

Table 13 summarizes critical values such as starting log-likelihood value and convergence log-likelihood value, and the number of iterations due to different step sizes. As shown in Table 13, calculated starting log-likelihood values are exactly the same for all four algorithms and for any step size used within each algorithm due to the condition applied in guessing of starting points, $\beta_k = [\beta_1 = 0, \beta_2 = 0, \beta_3 = 0, \dots, \beta_9 = 0]^T$. On the other hand, convergence values are slightly different among the step sizes only below one decimal point and so the difference can be ignored and all runs are considered to be converged at the same log-likelihood value.

The convergence behavior is identical with previous experiments, except the fact that slight differences occur all over the three performance measures. These experimental

TABLE 13: Log-likelihood values obtained from four algorithm (exp. 1-2) (CR = 10^{-6}).

Algorithms & log-likelihood values	Step size (λ)							
	1/32	1/16	1/8	1/4	1/2	1	1	1
DFP-(1)								
Starting	-	-869.0964727	-869.0964727	-869.0964727	-869.0964727	-	-	-
Convergence	-	-705.5400439	-705.5415497	-705.5435313	-705.5461356	-	-	-
Iteration num.	-	216	139	95	69	-	-	-
BFGS-(2)								
Starting	-869.0964727	-869.0964727	-869.0964727	-869.0964727	-869.0964727	-869.0964727	-869.0964727	-869.0964727
Convergence	-705.5385897	-705.5386006	-705.5386834	-705.5390061	-705.5385813	-705.5391969	-705.5391969	-705.5391969
Iteration num.	274	158	100	66	48	36	36	36
Algorithms & log-likelihood values								
	LL(0), UL(1)	LL(0), UL(2)	LL(0), UL(3)	LL(0), UL(4)	LL(0), UL(5)	LL(0), UL(6)	LL(0), UL(7)	LL(0), UL(8)
DFP-GSM-(3)								
Starting	-869.0964727	-869.0964727	-869.0964727	-869.0964727	-869.0964727	-869.0964727	-869.0964727	-869.0964727
Converge	-705.5458168	-705.5482967	-705.5482987	-705.5483539	-705.5494868	-705.5491385	-705.5485776	-705.5494614
Iteration num.	42	17	17	16	14	14	14	13
BFGS-GSM-(4)								
Starting	-869.0964727	-869.0964727	-869.0964727	-869.0964727	-869.0964727	-869.0964727	-869.0964727	-869.0964727
Converge	-705.5422923	-705.5485768	-705.5437015	-705.5494556	-705.5493428	-705.5498048	-705.5440254	-705.5440296
Iteration num.	26	15	16	13	13	12	14	14

results are clearly presented in the graph. Overall, it was observed that as the step size decreased, the slope became gradually gentle and the number of interaction increased.

As shown in Figure 8, the profile of graph at the last iteration part also is shorter than that of experiment 1-1 ($CR = 10^{-4}$), but as compared with experiments 1-2 ($CR = 10^{-4}$) and experiment 1-2 ($CR = 10^{-5}$), the overlapped part becomes much longer. It means more iteration is conducted for each algorithm in order to satisfy a new stringent criterion charged for stopping the iteration process. It also means that the number of iterations and convergence time can be adjusted, if desired.

4.2. Numerical Experiment 2. In experiment 2, the applied convergence criterion is $G_t'(-H_t)^{-1}G_t < 10^{-4}$. The main concern of this section is to recognize the importance of guessing starting points and exhibit different estimation performance using experimental estimation runs with different starting points. A new method or an idea of finding a good initial guess of the starting point is not the subject of this section. The point of this section is that it might be important just to know that changing initial starting points can make a difference in estimating performance and that the task of developing a new systematic way of guessing starting points that guarantee a better operational performance might be a useful future research topic. For example, refer to the research conducted by Liu and Mahmassani [33] for additional information on this subject. To shed a light on this new interesting topic, in this section, a new assumed vector of parameter applied to the four algorithms is written as $\beta_k = [\beta_1 = -0.01, \beta_2 = -0.1, \beta_3 = 0, \dots, \beta_9 = 0]'$. This vector of parameter is randomly chosen among the vector of parameters that work in estimation.

The results obtained from this experimental runs are shown in Tables 14 and 15. In addition, the detailed results are compared graphically in Figures 9 and 10 in terms of three performance measures or according to variances of log-likelihood values obtained for four algorithms and using an assumption on a vector of parameter.

It is useful to know that applying different vectors of starting points can cause diverse variations in estimation performances that can enhance the situation or make it worse and that trying to find an advanced way of guessing starting points guarantees improvement in the estimation performance in targeted performance measures. Therefore, this experiment could be recognized as a potential topic that a researcher can undertake.

Overall, the results obtained from experimental estimation are as follows. (1) Starting log-likelihood values are enhanced in its magnitude due to adopting new starting points. (2) In case of algorithms, the convergence time is greatly improved compared to the results of other experiments and the improvement in test results means that a researcher's prior knowledge on parameter to be estimated is correct and new guessed starting points make an improvement in estimation performance. (3) The convergence log-likelihood values are deviated between algorithms, but the reason is not clear here.

4.3. Numerical Experiment 3. In this section, we consider another way of guessing the initial Hessian matrix, which is assumed to be an identity matrix before starting estimation and continuously updated as the estimation processes in the DFP, BFGS and DFP-GSM, BFGS-GSM algorithms. Another way that is tried in this experiment is to incorporate the first iteration routine that is used to calculate the Hessian matrix in both BHHH and BHHH-2 algorithms instead of assuming identity matrix as a starting Hessian matrix.

By adding codes associated with the calculation of the initial Hessian matrix in BHHH and BHHH-2 algorithms to general estimation procedures of the above four algorithms, new estimation codes are completed in this experiment. They are named as DFP (BHHH), DFP (BHHH-2), BFGS (BHHH), BFGS (BHHH-2) and DFP-GSM (BHHH), DFP-GSM (BHHH-2), BFGS-GSM (BHHH), and BFGS-GSM (BHHH-2). It is important to note that the only difference between two sets of algorithms (e.g., DFP and DFP (BHHH)) is the way of obtaining the initial Hessian matrix. The first algorithm (DFP) uses an identity matrix as an its initial Hessian matrix, and the second algorithm (DFP (BHHH)) uses an advanced way presented in this experiment to get an initial Hessian matrix.

The applied convergence criterion is $G_t'(-H_t)^{-1}G_t < 10^{-4}$ and the starting vector of parameters is $\beta_k = [\beta_1 = 0, \beta_2 = 0, \beta_3 = 0, \dots, \beta_9 = 0]'$. The estimation results obtained from this experimental runs are shown in Tables 16 and 17. The detailed results are compared graphically in Figures 11–14 in terms of three performance measures, variances of log-likelihood value.

An important lesson being learned through a given series of experimental estimations is to know that applying a different way of guessing initial Hessian matrix can make a difference in estimation performance in terms of performance measures, which can be directed to better or worse performance. In the context of this research, the experimental estimations tried in this experiment generally give better results than that of experiment 1-2 ($CR = 10^{-4}$) for the given performance measures. It means that a better way of guessing the initial Hessian matrix, unlike the general procedures of using an identity matrix as an initial Hessian matrix, would improve the estimation performances in the targeted performance measures (i.e., convergence time). This issue can be another topic having potential to generate fruitful research results in the field of econometric modeling.

The following explanations of the results are compared to the results of the experiment 1-2 ($CR = 10^{-4}$), which is estimated under the same condition of experiment 3 except initial Hessian matrix. The results obtained from experimental estimation are as follows. (1) In terms of a single iteration time, there is no big difference between the runs using a general routine shown in Figure 3 and the runs using a new routine proposed in experiment 3 shown in Figures 11 and 12. However, when using a new proposed routine, the single iteration time becomes similar to other runs irrespective of using both different step sizes (in the case of DFP (BHHH), BFGS (BHHH) and DFP (BHHH-2), and BFGS (BHHH-2)) and different uncertain interval (in the case of DFP-GSM (BHHH), BFGS-GSM (BHHH) and

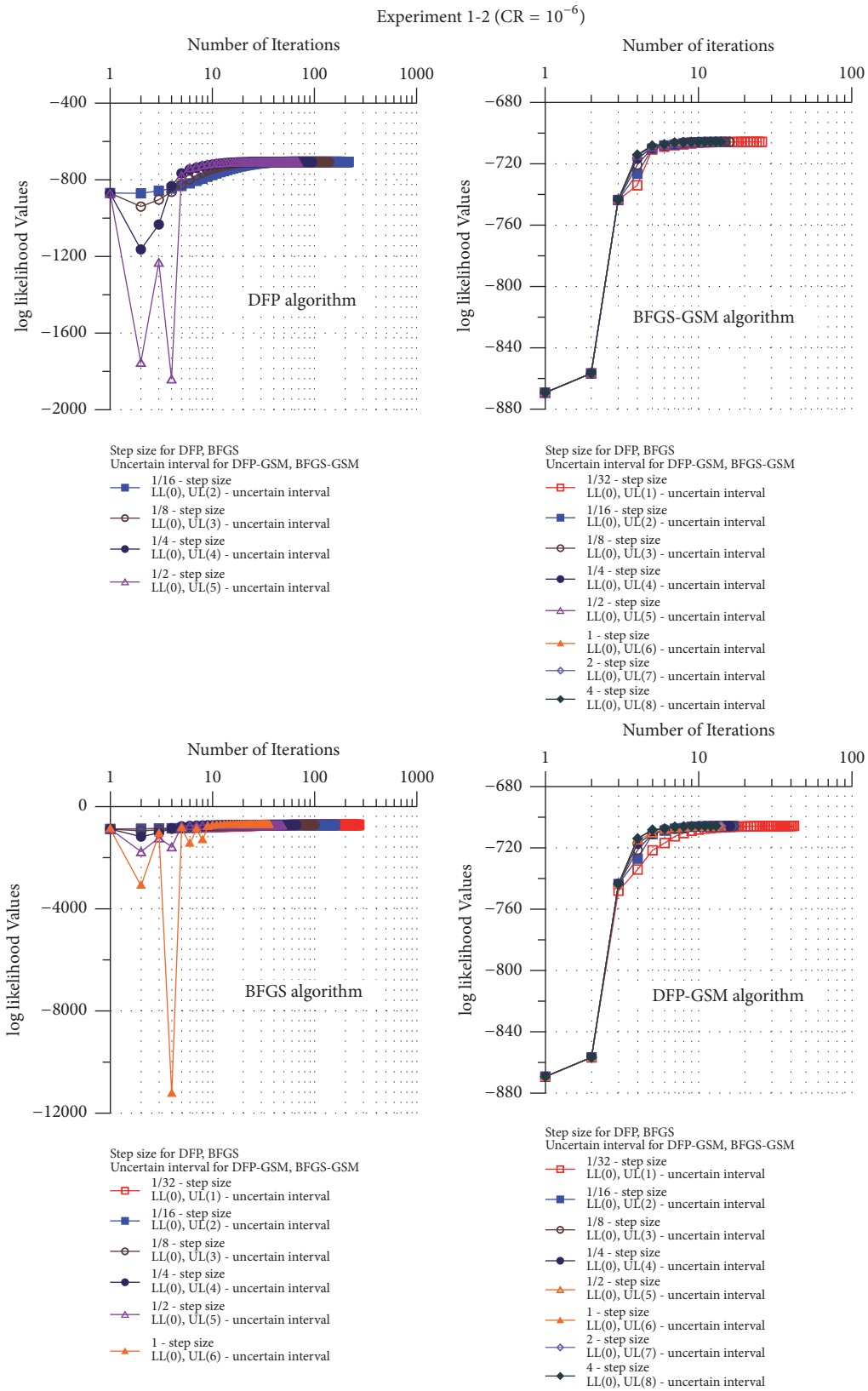


FIGURE 8: Variances of log-likelihood values for four algorithms (exp. 1-2) (CR = 10^{-6}).

TABLE 14: The results of four estimation algorithms (exp. 2) ($CR = 10^{-4}$).

Algorithms & performance measures	Step size (λ)							
	1/32	1/16	1/8	1/4	1/2	1	2	4
DFP-(1)								
One iteration time	-	0:00:19	0:00:18	0:00:17	0:00:18	-	-	-
Number of iterations	-	78	55	43	31	-	-	-
Convergence time	-	0:24:02	0:16:46	0:12:24	0:09:16	-	-	-
BFGS-(2)								
One iteration time	0:00:18	0:00:18	0:00:18	0:00:18	0:00:18	0:00:17	-	-
Number of iterations	148	77	48	32	21	16	-	-
Convergence time	0:43:40	0:22:52	0:13:59	0:09:19	0:06:11	0:04:24	-	-
Interval of uncertainty								
Algorithms & performance measures	LL(0), UL(1)	LL(0), UL(2)	LL(0), UL(3)	LL(0), UL(4)	LL(0), UL(5)	LL(0), UL(6)	LL(0), UL(7)	LL(0), UL(8)
DFP-GSM-(3)								
One iteration time	0:03:01	0:03:28	0:03:33	0:01:17	0:03:39	0:01:46	0:01:46	0:01:47
Number of iterations	23	11	9	10	10	9	8	8
Convergence time	1:09:07	0:38:08	0:31:47	0:12:42	0:36:23	0:15:50	0:14:05	0:14:12
BFGS-GSM-(4)								
One iteration time	0:01:13	0:01:21	0:01:22	0:01:17	0:01:17	0:01:21	0:01:22	0:01:26
Number of iterations	15	11	11	11	10	9	10	9
Convergence time	0:18:15	0:14:44	0:14:58	0:14:04	0:12:49	0:12:08	0:13:33	0:12:46

TABLE 15: Log-likelihood values obtained from four algorithm (exp. 2) ($CR = 10^{-4}$).

Algorithms & log-likelihood values	Step size (λ)							
	1/32	1/16	1/8	1/4	1/2	1	1	1
DFP-(1)								
Starting	-	-860.5995082	-860.5995082	-860.5995082	-860.5995082	-	-	-
Convergence	-	-705.5871328	-705.6177499	-705.6517527	-705.6940772	-	-	-
Iteration num.	-	78	55	43	31	-	-	-
BFGS-(2)								
Starting	-860.5995082	-860.5995082	-860.5995082	-860.5995082	-860.5995082	-860.5995082	-860.5995082	-860.5995082
Convergence	-705.5763454	-705.5817776	-705.5869734	-705.5896347	-705.6084271	-705.6838354	-705.6838354	-705.6838354
Iteration num.	148	77	48	32	21	16	16	16
Algorithms & log-likelihood values								
	LL(0), UL(1)	LL(0), UL(2)	LL(0), UL(3)	LL(0), UL(4)	LL(0), UL(5)	LL(0), UL(6)	LL(0), UL(7)	LL(0), UL(8)
DFP-GSM-(7)								
Starting	-860.5995082	-860.5995082	-860.5995082	-860.5995082	-860.5995082	-860.5995082	-860.5995082	-860.5995082
Convergence	-705.7088334	-705.7191719	-705.8555421	-705.7644033	-705.7724656	-705.7661276	-705.8789338	-705.8718341
Iteration num.	23	11	9	10	10	9	8	8
BFGS-GSM-(8)								
Starting	-860.5995082	-860.5995082	-860.5995082	-860.5995082	-860.5995082	-860.5995082	-860.5995082	-860.5995082
Convergence	-705.5878645	-705.6088839	-705.5924319	-705.5903149	-705.6002936	-705.6186549	-705.5948914	-705.6192299
Iteration num.	15	11	11	11	10	9	10	9

TABLE 16: The experimental results of four algorithms responding to a new way of calculating the initial Hessian matrix adopted from either BHHH or BHHH-2 algorithm (exp. 3) ($CR = 10^{-4}$).

Algorithms & performance measures	Step size (λ)							
	1/32	1/16	1/8	1/4	1/2	1	2	4
DFP-(BHHH)								
One iteration time	-	0:00:18	0:00:18	0:00:24	0:00:20	-	-	-
Number of iterations	-	74	37	19	11	-	-	-
Convergence time	-	0:22:27	0:11:05	0:07:20	0:03:31	-	-	-
DFP-(BHHH-2)								
One iteration time	-	0:00:20	0:00:20	0:00:20	0:00:19	-	-	-
Number of iterations	-	74	37	19	10	-	-	-
Convergence time	-	0:24:10	0:12:07	0:06:13	0:03:00	-	-	-
BFGS-(BHHH)								
One iteration time	0:00:18	0:00:17	0:00:18	0:00:18	0:00:20	0:00:17	-	-
Number of iterations	147	74	37	19	10	6	-	-
Convergence time	0:42:48	0:21:01	0:10:44	0:05:33	0:03:07	0:01:38	-	-
BFGS-(BHHH-2)								
One iteration time	0:00:18	0:00:18	0:00:18	0:00:18	0:00:19	0:00:15	-	-
Number of iterations	147	73	37	19	10	6	-	-
Convergence time	0:44:29	0:22:12	0:11:11	0:05:42	0:03:03	0:01:40	-	-
Algorithms & performance measures								
	LL(0), UL(1)	LL(0), UL(2)	LL(0), UL(3)	LL(0), UL(4)	LL(0), UL(5)	LL(0), UL(6)	LL(0), UL(7)	LL(0), UL(8)
DFP-GSM-(BHHH)								
One iteration time	0:02:40	0:01:34	0:01:36	0:01:41	0:01:41	0:01:47	0:01:47	0:01:47
Number of iterations	7	6	6	6	6	6	6	6
Convergence time	0:18:32	0:09:13	0:09:26	0:09:57	0:09:57	0:10:32	0:10:33	0:10:34
DFP-GSM-(BHHH-2)								
One iteration time	0:01:21	0:01:36	0:01:36	0:01:42	0:01:42	0:01:47	0:01:49	0:01:46
Number of iterations	7	6	6	6	6	6	6	6
Convergence time	0:09:19	0:09:25	0:09:25	0:10:01	0:10:05	0:10:31	0:10:43	0:10:28
BFGS-GSM-(BHHH)								
One iteration time	0:01:12	0:01:16	0:01:16	0:01:18	0:01:20	0:01:24	0:01:25	0:01:22
Number of iterations	5	6	6	6	6	6	6	6
Convergence time	0:07:08	0:07:28	0:07:26	0:07:37	0:07:52	0:08:17	0:08:21	0:08:05
BFGS-GSM-(BHHH-2)								
One iteration time	0:01:06	0:01:13	0:01:15	0:01:18	0:01:18	0:01:22	0:01:21	0:01:22
Number of iterations	6	6	6	6	6	6	6	6
Convergence time	0:6:46	0:07:30	0:07:38	0:08:00	0:07:54	0:08:19	0:08:17	0:08:32

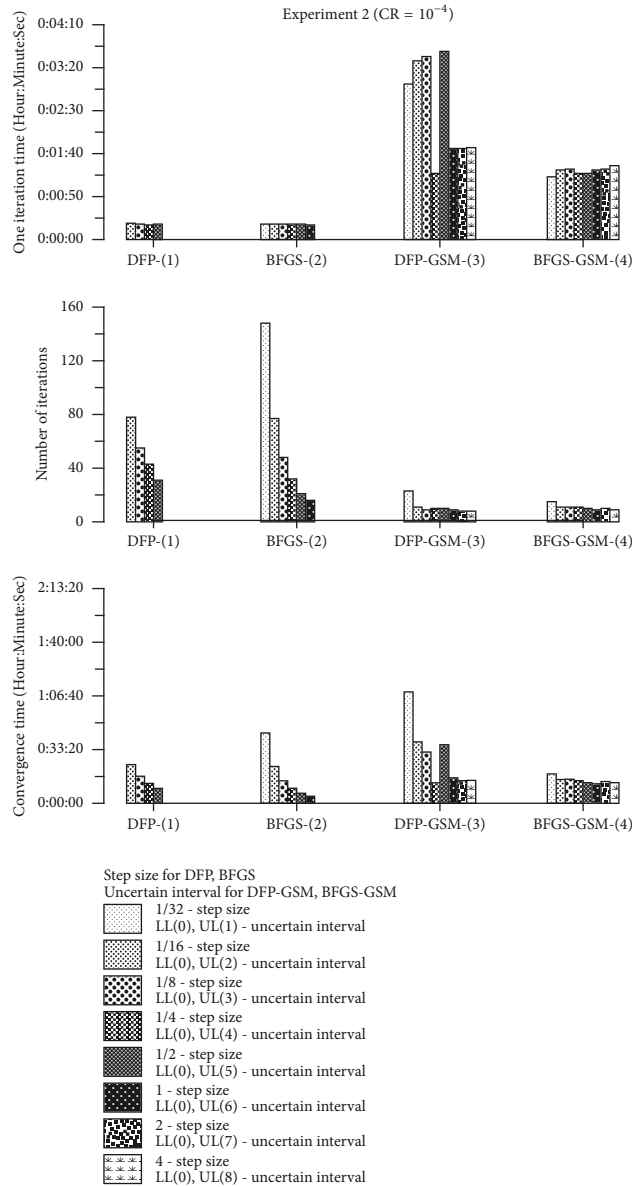


FIGURE 9: The results of estimation for four algorithms (exp. 2) (CR = 10⁻⁴).

DFP-GSM (BHHH-2), and BFGS-GSM (BHHH-2)). (2) In terms of the number of iterations, the routine proposed in experiment 3 shows a better performance. The number of iteration is decreased. (3) In terms of the convergence time, the big difference is observed between experiment 1-2 (CR = 10⁻⁴) and experiment 3. The experiment 3 takes much less convergence time than experiment 1-2 (CR = 10⁻⁴). The reason is that using a more accurate initial Hessian matrix at the initial stage of estimation, which can be calculated and input into the routine of estimation, instead of using an identity matrix would decrease the iteration time needed to reach the convergence Hessian matrix.

5. Summary and Discussion

Three important factors affecting strongly overall estimation performance in discrete choice modeling are detected

through extensive experimental estimation which is classified with three categorical experiment: convergence criterion, initial guessing of starting points, and initial Hessian matrix.

- (i) The first category considered in this research is to adopt convergence criterion that is different in both its mathematical expression (compare experiment 1-1 and experiment 1-2) and its degree of stopping precision (experiment 1-2) as given in Table 5. The purpose of these experiments was to show the evidence of changing estimation performance, to compare them in terms of performance measures, and to recognize the convergence criterion as an important factor dominating the performance of estimation process.

- (a) As an initial guess of the starting value of a vector of parameter, the value of $\beta_k = [\beta_1 = 0, \beta_2 = 0, \beta_3 = 0, \dots, \beta_9 = 0]^T$ is used for

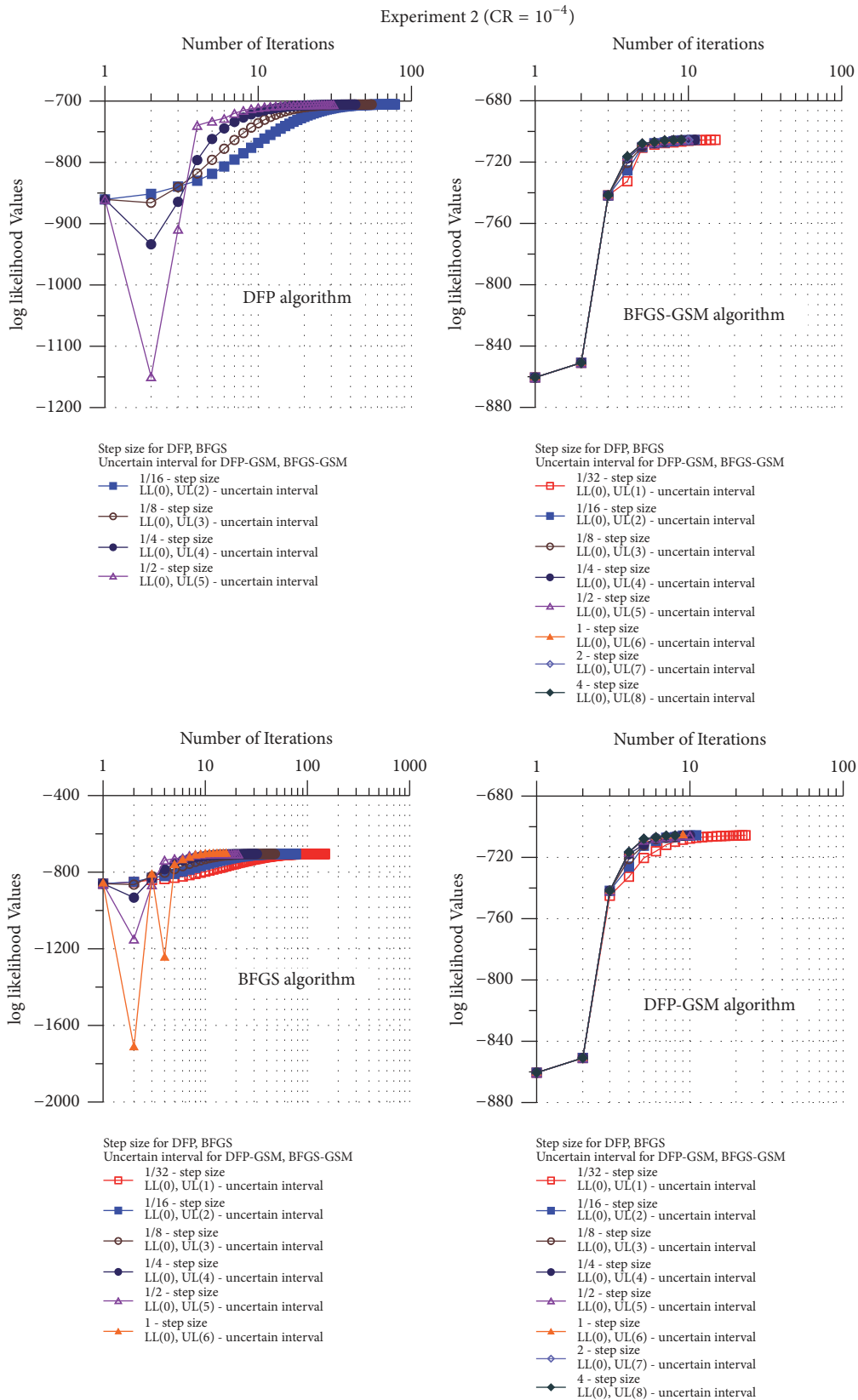


FIGURE 10: Variances of log-likelihood values for four algorithms (exp. 2) (CR = 10⁻⁴).

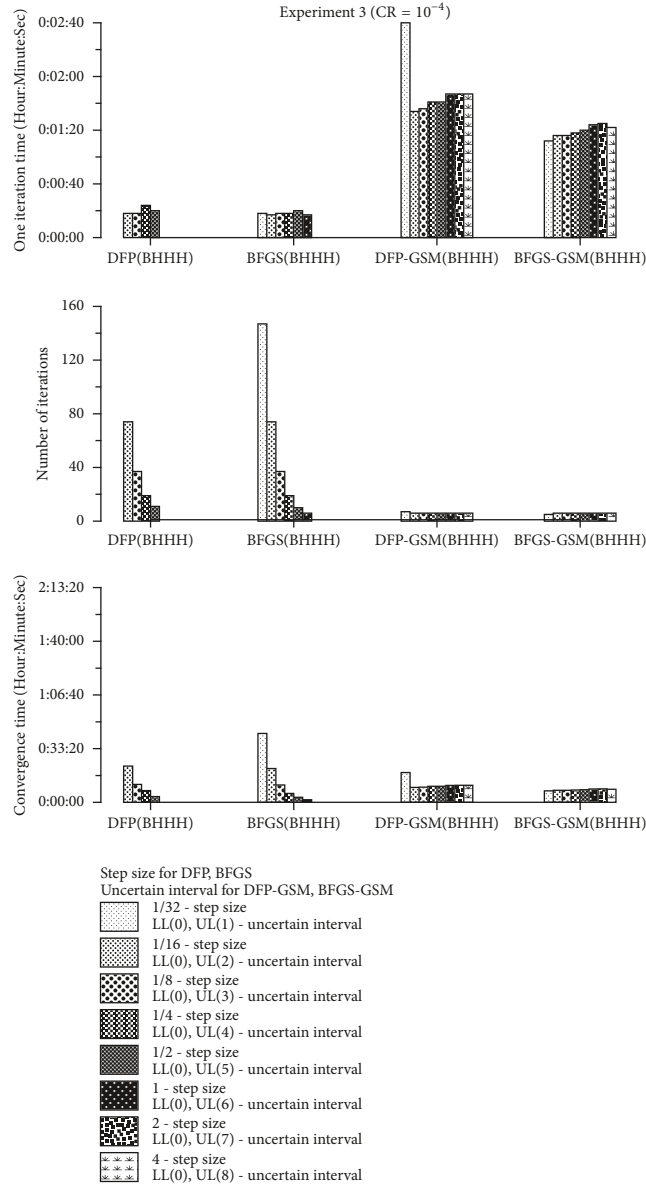


FIGURE 11: The results of four algorithms responding to a new way of calculating the initial Hessian matrix adopted from BHHH algorithm (exp. 3) (CR = 10⁻⁴).

- all experiments included in the first categorical experiment.
- (b) The first mathematical convergence criterion is most widely used as a stopping criterion. It is adopted from Ben-Akiva and Lermans [1] and it has a mathematical expression, $[(1/k) \sum_{k=1}^k (\beta_{t+1,k} - \beta_{t,k})^2]^{1/2} < 10^{-4}$.
- (c) The second convergence criterion has the following common mathematical expression: $G'_t(-H_t)^{-1}G_t < CR$. It was suggested by Besley [17]. In this research, by changing a CR gradually to other more stringent values such as 10⁻⁴, 10⁻⁵, and 10⁻⁶, we were trying to show that

- model estimation performance is sensitive in changing of CR.
- (d) The first two algorithms (i.e., ① DFP, ② BFGS) are estimated using various step sizes throughout all experiments and only the converged runs are summarized and compared.
- (e) The last two algorithms (i.e., ③ DFP-GSM, ④ BFGS-GSM) use the golden section line search (GSM) method instead of choosing a step size manually before the estimation procedure begins. The GSM method finds an optimal value of step size automatically during iteration, with which the log-likelihood function is maximized in every iteration process.

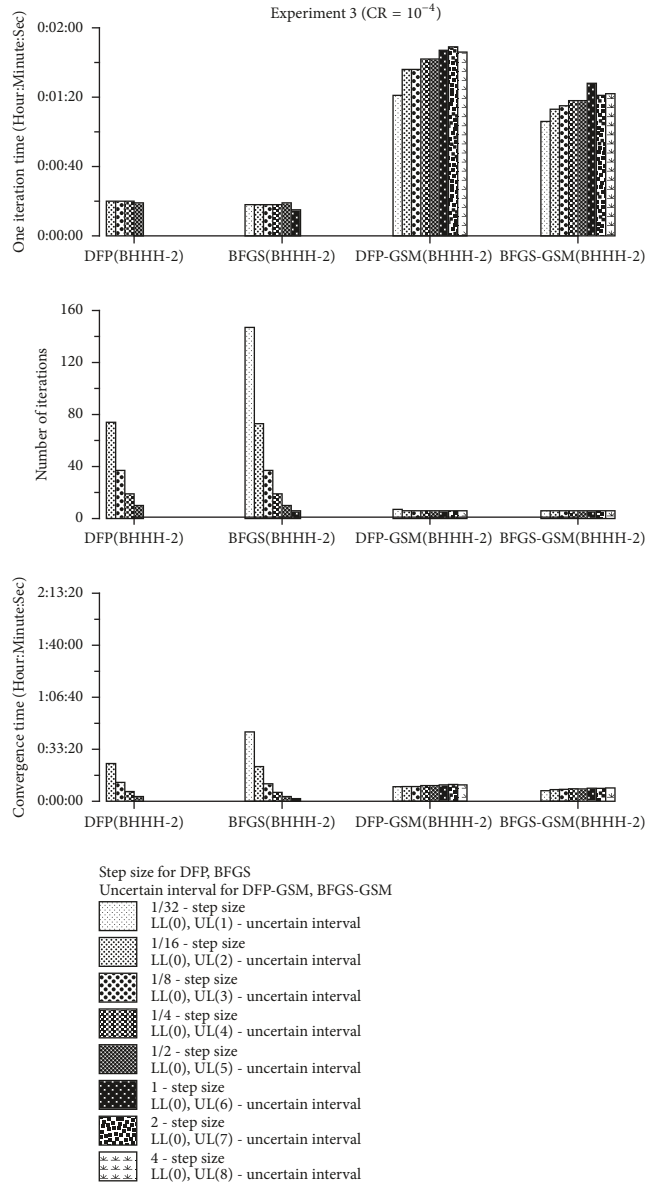


FIGURE 12: The results of four algorithms responding to a new way of calculating the initial Hessian matrix adopted from BHHH-2 algorithm (exp. 3) (CR = 10⁻⁴).

- (f) To use the GSM in line search work, we should assign the interval of uncertainty, in which the optimal step size is investigated automatically and it is used in every iteration estimation procedure. In the context of this research, eight different uncertain intervals are proposed and the experimental estimation is run using each of them. Typically, there are no limitations in choosing uncertain intervals, if a researcher chooses the interval having a distance of between 0 and a certain value over 0 [31].
- (g) From the results of the first categorical experiment, we identify several facts as follows. (1) In spite of using the same model specification

(i.e., using the same multinomial logit choice function and the same utility functions for each alternative) and common data set, there exist large or small differences among four estimations algorithms in terms of three performance measures. (2) Depending on the step size, large variations are generated in the estimation performance among four different algorithms or even within the same algorithm. In conclusion, the step size chosen by a researcher at an initial stage of model estimation can be an important control factor controlling the entire estimation performance, while the model is under estimation. (3) As opposed to choosing a step size randomly as required in the first

Experiment 3 (CR = 10^{-4})

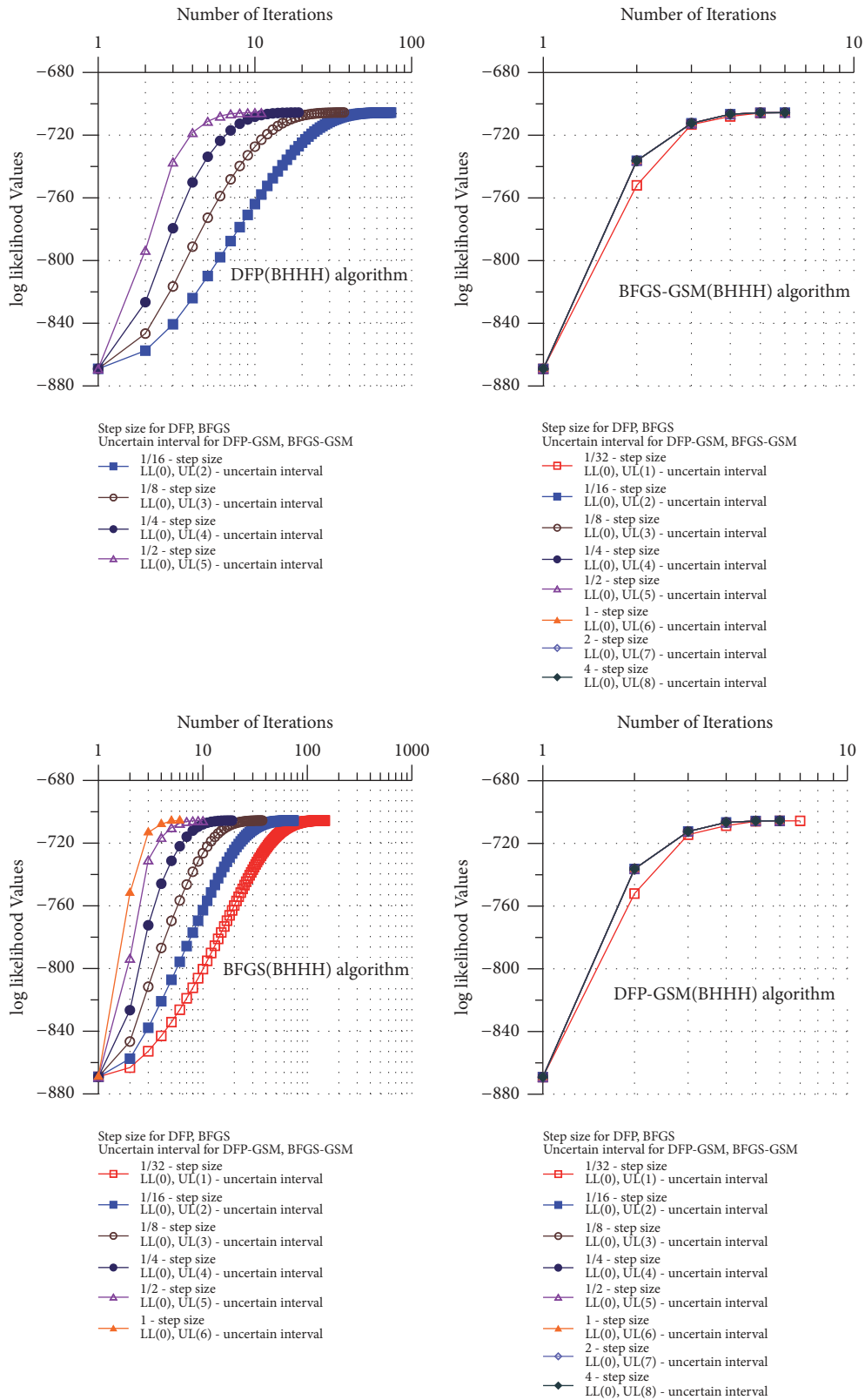


FIGURE 13: Variances of log-likelihood values for four algorithms responding to a new way of calculating the initial Hessian matrix adopted from BHHH algorithm (exp. 3) (CR = 10^{-4}).

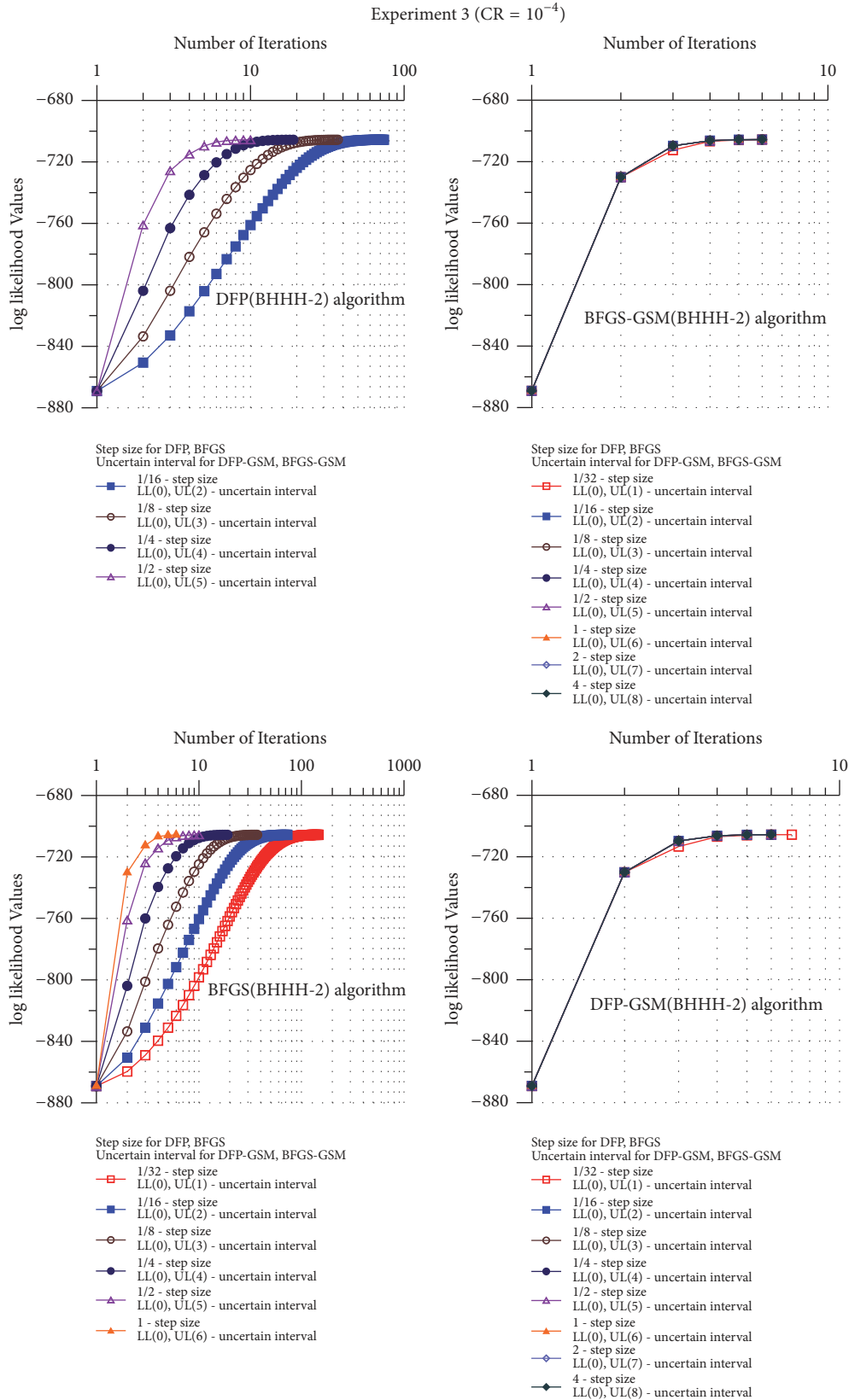


FIGURE 14: Variances of log-likelihood values for four algorithms responding to a new way of calculating the initial Hessian matrix adopted from BHHH-2 algorithm (exp. 3) (CR = 10^{-4}).

two estimation algorithms, in this research, a new algorithm employs a golden section line search method (GSM) for finding the optimal step size automatically in every iteration. The two algorithms tried, DFP-GSM and BFGS-GSM, show a stable estimation performance in terms of measures used regardless of using any uncertain intervals. By using this method, a researcher can minimize the risk of choosing a bad step size, which can happen in the case of first two estimation algorithms.

- (h) The facts observed from a comparison of experiment 1-1 (CR = 10^{-4}) with experiment 1-2 (CR = 10^{-4}) are as follows. (1) The result of experiment 1-2 (CR = 10^{-4}) shows a better estimation performance in terms of three performance measures than that of experiment 1-1 (CR = 10^{-4}): less time in one iteration, smaller number of iterations, and faster convergence time. (2) This means that different types of stopping criteria can lead to different estimation results and show a different estimation performance, even in the same modeling situation.
- (i) Observations drawn from the comparisons of experiments 1-2 (CR = 10^{-4}), experiment (CR = 10^{-5}), and experiment (CR = 10^{-6}) are as follows. In this case, the same mathematical expression of stopping criterion is used but different stopping precision is employed. (1) In the results obtained with increasing precision of stopping criterion, the performance degenerates in terms of performance measures: taking more time in one iteration, more iterations before convergence, and taking more time before convergence. (2) The reason for this phenomenon can be attributed to the fact that applying the more stringent stopping criterion necessitates more calculation time to satisfy the more stringent stopping criterion before terminating iteration procedures. (3) This result implies the fact that a different stopping precision applied in the same mathematical expression of stopping criterion is another control measure having a big impact on model estimation performance.
- (ii) The second category is to use a different initial guess of the parameter estimates of model. This is the subject of experiment 2.
- (a) Here, the stopping criterion is $G_t'(-H_t)^{-1}G_t < 10^{-4}$.
- (b) The main concern of experiment 2 is to recognize the importance of initial guess of starting points and to show an example of experimental estimation results reflecting the fact that the estimation performance can be affected by use of different starting points.
- (c) The vector of parameter is $\beta_k = [\beta_1 = -0.01, \beta_2 = -0.1, \beta_3 = 0, \dots, \beta_9 = 0]'$.

- (d) It is important to know that changing initial starting points can result in different estimation performance and that the development of a new systematic way of guessing starting points guaranteeing a better operational performance would be a useful future research topic.

Overall, the results obtained from experimental estimation are as follows. (1) Starting log-likelihood values are different from other categorical experiments due to adopting new sets of starting points. (2) Different convergence time compared to experiment 1-2 (CR = 10^{-4}) is examined. This means that if a researcher can specify properly parameter's characteristics based on good prior knowledge (e.g., sign or magnitude of parameter estimates) or if a researcher can develop a new scientific way of guessing starting points, estimation performance may be improved.

- (iii) The third experimental category is on employing a different way of calculating the initial Hessian matrix instead of using identity matrix. This experiment is carried out in experiment 3.
- (a) The applied convergence criterion is $G_t'(-H_t)^{-1}G_t < 10^{-4}$.
- (b) The applied initial guess of starting points of a vector of parameters is $\beta_k = [\beta_1 = 0, \beta_2 = 0, \beta_3 = 0, \dots, \beta_9 = 0]'$.
- (c) All four algorithms are used for experiment 3 which use an identity matrix as an initial Hessian matrix in the model estimation procedures (i.e., DFP, BFGS, DFP-GSM, and BFGS-GSM).
- (d) An experimented way of getting a new initial Hessian matrix is possible by incorporating the method of both BHHH and BHHH-2 for getting an initial Hessian matrix in the first iteration.
- (e) By adding computer codes associated with the calculation of the initial Hessian matrix used in both BHHH and BHHH-2 algorithms to general estimation procedures of the above four algorithms, new estimation codes that differ only in the way of getting initial Hessian matrix were written. They are named as DFP (BHHH), BFGS (BHHH), DFP (BHHH-2), BFGS (BHHH-2) and DFP-GSM (BHHH), BFGS-GSM (BHHH), DFP-GSM (BHHH-2), and BFGS-GSM (BHHH-2).
- (f) An important lesson educated from a given series of experimental estimations is to know that applying a different way of guessing initial Hessian matrix can make a difference in estimation performance assessed according to specified performance measures.
- (g) In the context of this research, the results of experiment 3 are in general better than that of experiment 1-2 (CR = 10^{-4}). These use the same condition of stopping criterion but differ in terms of the method for acquiring the initial

Hessian matrix. It means that a better way of guessing the initial Hessian matrix, unlike the general procedures (i.e., an identity matrix as an initial Hessian matrix), would improve the estimation performance such as convergence time or the number of iterations. This issue can be another topic with potential to generate useful research results in econometric modeling.

- (h) The results obtained from experiment 3 are as follows. (1) In terms of the single iteration time, there is no big difference between experiment 1-2 ($CR = 10^{-4}$) and experiment 3. However, when using a new routine for experiment 3, the single iteration time becomes similar among runs irrespective of using both different step size (in the case of DFP (BHHH), BFGS (BHHH) and DFP (BHHH-2), and BFGS (BHHH-2)) and different uncertain intervals (in the case of DFP-GSM (BHHH), BFGS-GSM (BHHH) and DFP-GSM (BHHH-2), and BFGS-GSM (BHHH-2)). (2) As for the number of iterations, the new routine used in experiment 3 shows better performance (i.e., a decrease in the number of iterations). (3) In terms of the convergence time, there is a big difference between experiment 1-2 ($CR = 10^{-4}$) and experiment 3. The experiment 3 takes much less convergence time than that of experiment 1-2 ($CR = 10^{-4}$). (4) The reason for this phenomenon is that using a more advanced way of assuming initial Hessian matrix instead of using an identity matrix would decrease the iteration time needed to reach the final Hessian matrix of when the estimation is terminated.

6. Conclusions

There are several open sources for discrete choice modeling work, but none of these open source programs or packages provides an opportunity for researchers to go beyond the stage of writhing a few command lines. In many cases of using software packages, researcher can choose algorithms in very simple manner and run a program with a few lines of commands and obtain analysis results. However, there is nothing more to do or learn than simply making a choice for estimation without knowing the inner workings procedures of the chosen algorithm. The software packages do not provide information about the algorithms and their capabilities. Consequently, researchers are not in a position to obtain any information about how an algorithm works and what makes the difference between algorithms that they can choose from a tray of algorithms. These are the drawbacks of traditional modeling practices, which prevent transportation analysts/researchers being more exposed to numerical operational aspect of discrete choice model estimation ahead of convenience.

Understanding the calculation mechanism for pedagogically important estimation algorithms means that modelers could incorporate their ideas interpreted with the help of

computer codes into the estimation procedures. This is possible only when modelers are well informed on the theoretical aspects of estimation algorithms and the numerical operation procedure of how it works. Based on the theoretical information, for example, modeler might know where a new line search method is to be placed to find an optimal step size in order to enhance algorithms' performance. Other positive sides of being familiar with theoretical aspects of algorithms are that many numerical experiments could be freely tried and these efforts could end up with newly advanced estimation algorithms. Understanding the theoretical part of estimation algorithms and being able to write custom-built compute codes could be an asset for modelers to address the modeling challenges in more advanced way. This paper contributes in this regard by showing a practical application of nonlinear technique into the traditional estimation routines and thus demonstrating an improved performance of proposed algorithms compared to currently circulated estimation algorithms in the market.

The multinomial logit choice model is estimated in the environment of custom-built modeling. Quasi-Newton algorithm (i.e., DFP and BFGS) is coded for the real choice data. Moreover, extended Quasi-Newton algorithm (i.e., DFP-GSM and BFGS-GSM) is proposed for the first time in this research and its estimation performance is relatively compared to that of both DFP and BFGS. Quasi-Newton is popular due to its high rate of success of convergence in discrete choice model [30]. Therefore, it is crucial for a researcher to be able to conduct modeling with his/her own tailor-made code for this widely used algorithm. Furthermore, the understanding of estimation mechanism of this algorithm and handling it freely in the middle of model developing in discrete choice modeling are an asset to a researcher. From the experience gained in model estimation involving four different algorithms and the maximum log-likelihood estimator, the following conclusions are drawn.

- (1) DFP and BFGS show similar performance due to the similarity in calculating and updating Hessian matrix. Updating rule of Hessian matrix of this Quasi-Newton method differentiates these two algorithms from other algorithms. BFGS is slightly more complex in Hessian updating formula than that of DFP.
- (2) By incorporating the golden section line search method (GSM), two new estimation algorithms, called DFP-GSM and BFGS-GSM, are developed and tested for the first time in this research. These perform better than DFP and BFGS.
- (3) Important factors identified through experimental estimations are as follows:
 - (i) Selecting a step size is an important element when using the custom-built code. This research provides insights about this issue.
 - (ii) Convergence criterion is another important factor dominating the entire estimation performance.
 - (iii) Initial Hessian matrix is one of important factors and a new method for calculation of the

initial Hessian matrix used in this research proved to be beneficial.

- (4) Understanding estimation algorithms is critical in cases where a researcher develops a new advanced model that is not supported by commercial statistical packages. Specifically, in the current new generation of discrete choice modeling activities, these technique and knowledge of coding of given modeling situation are essential.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

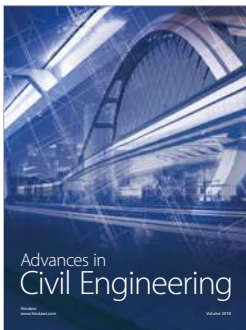
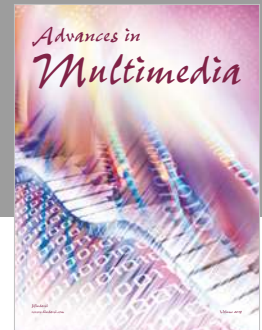
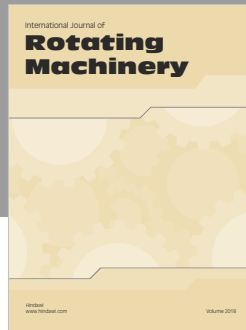
Acknowledgments

The authors are grateful to the Korea Transport Institute for providing the revealed preference data utilized for the numerical experiments in the study.

References

- [1] M. Ben-Akiva and S. R. Lermans, "Discrete Choice Analysis Theory and Application to Travel Demand," 1985, MIT, MA.
- [2] V. Psaraki and C. Abacoumkin, "Access mode choice for relocated airports: the new athens International Airport," *Journal of Air Transport Management*, vol. 8, no. 2, pp. 89–98, 2002.
- [3] H.-J. Roh, "Mode choice behavior of various airport user groups for ground airport access," *Open Transportation Journal*, vol. 7, no. 1, pp. 43–55, 2013.
- [4] M. L. Tam, W. H. K. Lam, and H. P. Lo, "Incorporating passenger perceived service quality in airport ground access mode choice model," *Transportmetrica*, vol. 6, no. 1, pp. 3–17, 2009.
- [5] S. Choo, S. You, and H. Lee, "Exploring characteristics of airport access mode choice model," *Transportation Planning and Technology*, vol. 36, no. 4, 2013.
- [6] C. Chen, J. Ma, Y. Susilo, Y. Liu, and M. Wang, "The promises of big data and small data for travel behavior (aka human mobility) analysis," *Transportation Research Part C: Emerging Technologies*, vol. 68, pp. 285–299, 2016.
- [7] G. Harvey, "Study of airport access mode choice," *Journal of Transportation Engineering*, vol. 112, no. 5, pp. 525–545, 1986.
- [8] C. R. Bhat, "Accommodating flexible substitution patterns in multi-dimensional choice modeling: formulation and application to travel mode and departure time choice," *Transportation Research Part B: Methodological*, vol. 32, no. 7, pp. 455–466, 1998.
- [9] C. Evertse and H. G. Visser, "Real-time airport surface movement planning: Minimizing aircraft emissions," *Transportation Research Part C: Emerging Technologies*, vol. 79, pp. 224–241, 2017.
- [10] J. D. Hunt and J. E. Abraham, "Stated response examination of factors influencing commercial movement route choice behavior," *Journal of Transportation Engineering*, pp. 72–77, 2004.
- [11] H. Ohashi, T.-S. Kim, T. H. Oum, and C. Yu, "Choice of air cargo transshipment airport: An application to air cargo traffic to/from Northeast Asia," *Journal of Air Transport Management*, vol. 11, no. 3, pp. 149–159, 2005.
- [12] M. R. Tayyaran and A. M. Khan, "Telecommuting and residential location decisions: Combined stated and revealed preferences model," *Canadian Journal of Civil Engineering*, vol. 34, no. 10, pp. 1324–1333, 2007.
- [13] Y. Dong, S. Wang, L. Li, and Z. Zhang, "An empirical study on travel patterns of internet based ride-sharing," *Transportation Research Part C: Emerging Technologies*, vol. 86, pp. 1–22, 2018.
- [14] G. Harvey, "Airport choice in a multiple airport region," *Transportation Research Part A: General*, vol. 21, no. 6, pp. 439–449, 1987.
- [15] S. Hess and J. W. Polak, "Mixed logit modelling of airport choice in multi-airport regions," *Journal of Air Transport Management*, vol. 11, no. 2, pp. 59–68, 2005.
- [16] F. El Zarwi, A. Vij, and J. L. Walker, "A discrete choice framework for modeling and forecasting the adoption and diffusion of new transportation services," *Transportation Research Part C: Emerging Technologies*, vol. 79, pp. 207–223, 2017.
- [17] D. A. Besley, "On the computational competitiveness of full-information maximum-likelihood and three-stage least-squares in the estimation of nonlinear, simultaneous-equations models," *Journal of Econometrics*, vol. 9, no. 3, pp. 315–342, 1979.
- [18] D. A. Belsley, "On the efficient computation of the nonlinear full-information maximum-likelihood estimator," *Journal of Econometrics*, vol. 14, no. 2, pp. 203–225, 1980.
- [19] D. S. Bunch, "A comparison of algorithms for maximum likelihood estimation of choice models," *Journal of Econometrics*, vol. 38, no. 1-2, pp. 145–167, 1988.
- [20] D. Brownstone, D. S. Bunch, and K. Train, "Joint mixed logit models of stated and revealed preferences for alternative-fuel vehicles," *Transportation Research Part B: Methodological*, vol. 34, no. 5, pp. 315–338, 2000.
- [21] G. Li and A. Balakrishnan, "Models and algorithms for network reduction," *European Journal of Operational Research*, vol. 248, no. 3, pp. 930–942, 2016.
- [22] Q. Fan, X. Yan, and Y. Zhang, "Auto-selection mechanism of differential evolution algorithm variants and its application," *European Journal of Operational Research*, 2017.
- [23] G. Karakaya, M. Koksalan, and S. D. Ahipasaoglu, "Interactive algorithms for a broad underlying family of preference functions," *European Journal of Operational Research*, vol. 265, no. 1, pp. 248–262, 2018.
- [24] H.-J. Roh, P. K. Sahu, A. M. Khan, and S. Sharma, "Relative efficiency appraisal of discrete choice modeling algorithms using small-scale maximum likelihood estimator through empirically tailored computing environment," *Journal of Modern Transportation*, vol. 23, no. 1, pp. 67–79, 2015.
- [25] H.-J. Roh and A. M. Khan, "Enhancing algorithmic base for discrete choice modelling," *KSCSE Journal of Civil Engineering*, vol. 17, no. 7, pp. 1798–1809, 2013.
- [26] B. Lee Hyeong, *Hangeul Excel 2000*, Cyber Press, Seoul, Republic of Korea, 2015.
- [27] B. Lee Hyeong, *Hangeul Excel 2000 VBA*, Cyber Press, Seoul, Republic of Korea, 2016.
- [28] B. Lee Hyeong, *Visual Basic 6*, Cyber Press, Seoul, Republic of Korea, 2013.
- [29] H. Greene William, *Econometric Analysis*, Prentice Hall, Upper Saddle River, NJ, USA, 4th edition, 2000.
- [30] K. E. Train, *Discrete choice methods with simulation*, Cambridge University Press, Cambridge, Second edition, 2009.
- [31] S. Bazaraa Mokhtar, D. Hanif, and C. M., *Nonlinear Programming: Theory and Algorithms*, John Wiley & Sons, New Yourk, NY, USA, 1999.
- [32] D. McFadden, "Conditional logit analysis of qualitative choice behaviour," in *Frontiers in Econometrics*, P. Zarembka, Ed., pp. 105–142, Academic Press, New York, NY, USA, 1974.

- [33] Y.-H. Liu and H. S. Mahmassani, "Global maximum likelihood estimation procedure for multinomial probit (MNP) model parameters," *Transportation Research Part B: Methodological*, vol. 34, no. 5, pp. 419–449, 2000.



Hindawi

Submit your manuscripts at
www.hindawi.com

