Web Appendix for "A Choice Model of Utility Maximization and Regret Minimization"

Taegyu Hur

Greg M. Allenby

Table of Contents

Web Appendix A: DERIVATION OF REGRET FUNCTION

If we allow for heteroskedastic errors across different alternatives, we can assume that

$$\xi_j \sim N(0, \sigma_j^2)$$

$$\begin{aligned}
\mathbb{E}_\xi\Big[\max_{n\neq j}\{V_n+\eta_n+\xi_n\}\Big] &= \sum_{n\neq j}\Bigg[\int (V_n+\eta_n+\xi_n)f(\xi_n)\cdot Pr(V_n+\eta_n+\xi_n > V_m+\eta_m+\xi_m, \forall m\neq n)d\xi_n\Bigg] \\
&= \sum_{n\neq j}\Bigg[\int (V_n+\eta_n+\xi_n)f(\xi_n)F_{\xi_1}(V_n+\eta_n+\xi_n-V_1-\eta_1)\cdots \\
&\quad F_{\xi_J}(V_n+\eta_n+\xi_n-V_J-\eta_J)d\xi_n\Bigg] \\
&= \sum_{n\neq j}\Bigg[\int (V_n+\eta_n+\xi_n)f(\xi_n)\prod_{m\neq n}\Phi\left(\frac{V_n+\eta_n+\xi_n-V_m-\eta_m}{\sigma_m}\right)d\xi_n\Bigg]
\end{aligned}$$

Therefore,

$$\begin{aligned}
R_j &= \mathbb{E}_\xi\Big[\max_{n\neq j}\{V_n+\eta_n+\xi_n\}\Big]-\mathbb{E}_\xi\big[V_j+\eta_j+\xi_j\big] \\
&= \sum_{n\neq j}\Bigg[\int (V_n+\eta_n+\xi_n)f(\xi_n)\prod_{m\neq n}\Phi\left(\frac{V_n+\eta_n+\xi_n-V_m-\eta_m}{\sigma_m}\right)d\xi_n\Bigg]-V_j-\eta_j
\end{aligned}$$

Web Appendix B: DUAL-GOAL PROGRAMMING SOLUTIONS FOR DISCRETE CHOICE

We consider the simplest discrete choice situations where there are two goods to illustrate alternative cases of the original and ε-constraint formulations.

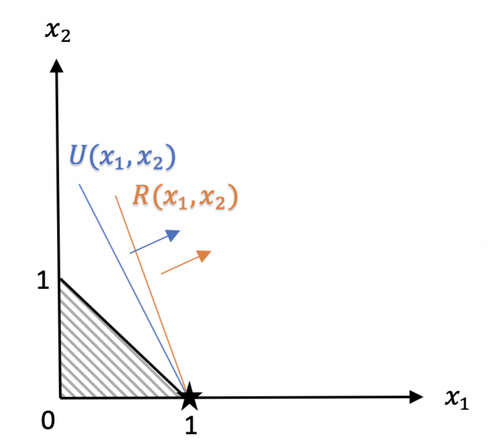
B.1 Case 1

Original problem.

$$\max \ U(x_1, x_2) = 2x_1 + x_2$$
$$\min \ R(x_1, x_2) = -3x_1 - x_2$$
$$s.t. \quad x_1 + x_2 \leq 1, \quad x_1, x_2 \in \{0, 1\}$$

The only solution to the problem is $(x_1^*, x_2^*) = (1, 0)$ and it is Pareto optimal since both objectives are optimized simultaneously at this point (Figure W1).

Figure W1: Dual-goal Programming Solution (Case 1)

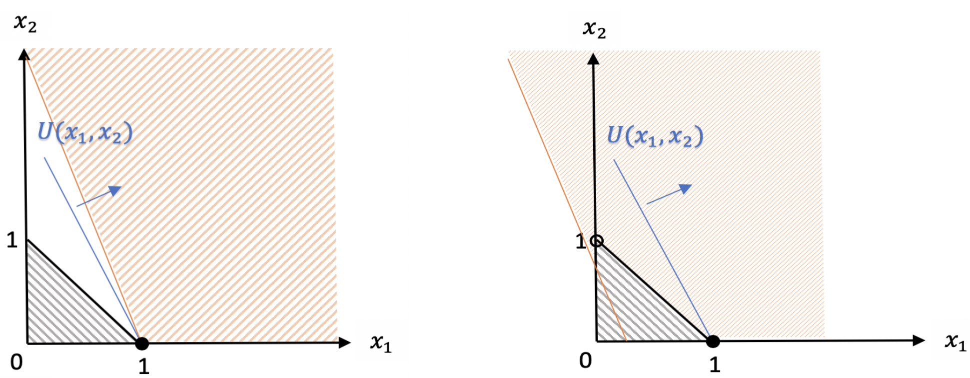


Corresponding ε-constraint formulations.

There are two corresponding ε-constraint formulations. We estimate both of the models and the model fits are different depending on which objective function is used as an additional constraint. The additional bound θ that determines the considered choice set of offerings is identified by the observed consideration set data in our proposed model.

Figure W2 depicts the first approach where R is used as an additional constraint. The solution derived from the ε-constraint approach is equivalent to the one from the original dual-goal programming as long as $\min R(x_1, x_2) \leq \theta$ (Figure W2(a) and W2(b)). Furthermore, if $\max R(x_1, x_2) \leq \theta$ (Figure W2(b)), we can see a special case where good 1 and 2 are both feasible and 1 is most preferred.

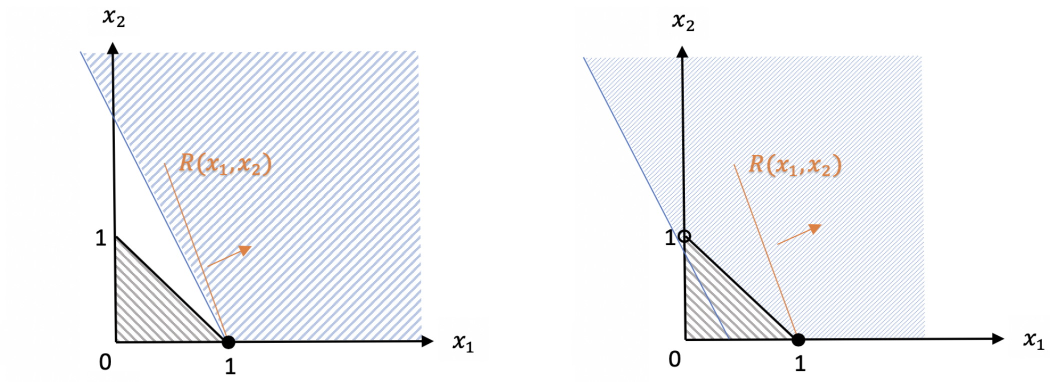
Figure W2: Additional constraint $R(x_1, x_2) \leq \theta_R$ (Case 1)



(a) (b)

Figure W3 is the case where utility function is used as an additional constraint (i.e., $-U(x_1, x_2) \leq \theta$). As long as $\min -U(x_1, x_2) \leq \theta$, the solution derived from ε-constraint approach is equivalent to the one from the original dual-goal problem (Figure W3(a) and W3(b)).

Figure W3: Additional constraint $-U(x_1, x_2) \leq \theta_U$ (Case 1)



(a) (b)

B.2 Case 2

We change the slope of the R function to illustrate the robustness of the ε-constraint formulation.

Original problem.

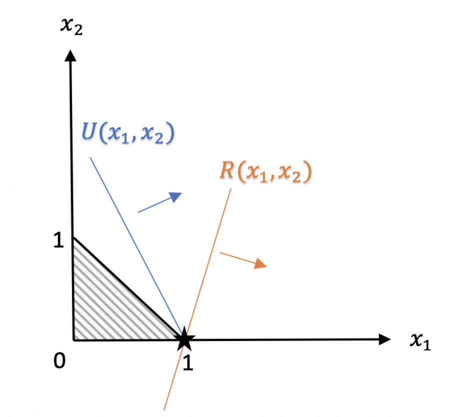
$$\max \; U(x_1, x_2) = 2x_1 + x_2$$

$$\min \; R(x_1, x_2) = -3x_1 + x_2$$

$$s.t. \quad x_1 + x_2 \leq 1, \quad x_1, x_2 \in \{0, 1\}$$

The only solution to the problem is $(x_1^*, x_2^*) = (1, 0)$ and it is Pareto optimal, or not dominated by other solution points. Both objectives are optimized simultaneously at this point (Figure W4).

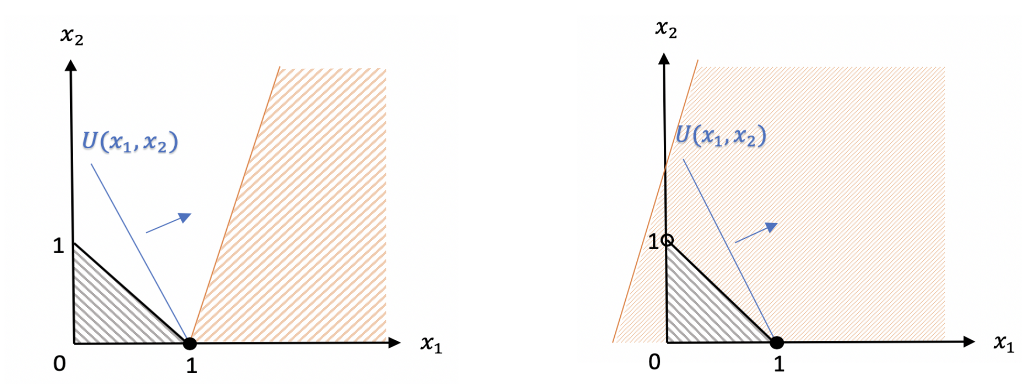
Figure W4: Dual-goal Programming Solution (Case 2)



Corresponding ε-constraint formulations.

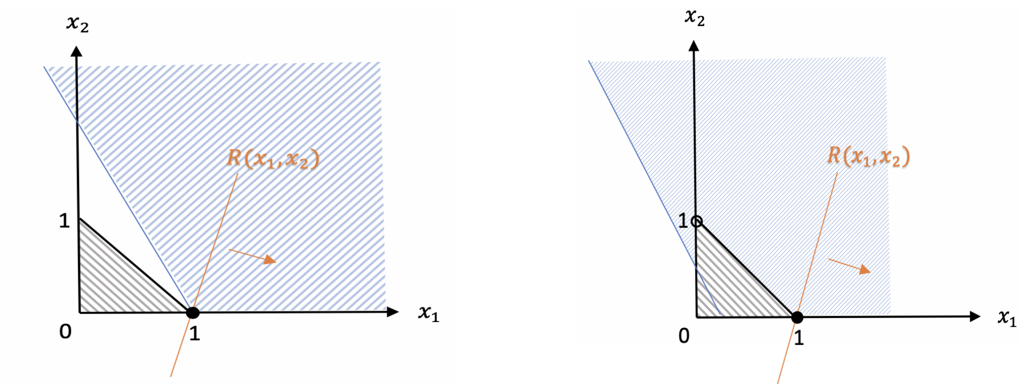
Using the same logic as in Case 1, we can draw the Figure W5 and W6, respectively. Both formulations lead to the same optimal solution.

Figure W5: Additional constraint $R(x_1, x_2) \leq \theta_R$ (Case 2)



(a) (b)

Figure W6: Additional constraint $-U(x_1, x_2) \leq \theta_U$ (Case 2)



(a) (b)

B.3 Case 3

We again change the slope of the R function to illustrate the case where there are multiple Pareto optimal solutions that are non-dominated.

Original problem.

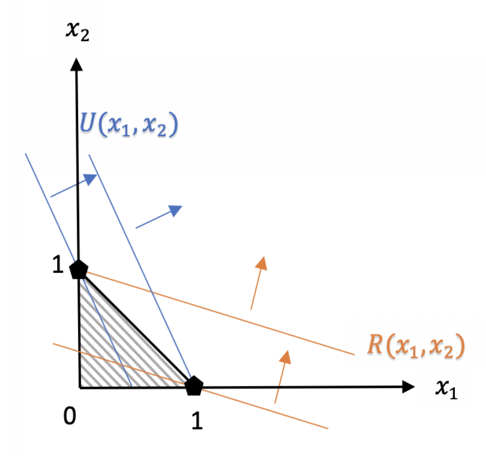
$$\max\ U(x_1, x_2) = 2x_1 + x_2$$

$$\min\ R(x_1, x_2) = -x_1 - 3x_2$$

$$s.t. \quad x_1 + x_2 \leq 1, \quad x_1, x_2 \in \{0, 1\}$$

Two possible solutions are $(x_1^*, x_2^*) = (1, 0)$ and $(x_1^*, x_2^*) = (0, 1)$. Both of them are Pareto optimal (Figure W7). For solution set $(x_1^*, x_2^*) = (1, 0)$, there is no other solution that better maximizes U and minimizes R simultaneously. The same claim can be made for the solution $(x_1^*, x_2^*) = (0, 1)$.

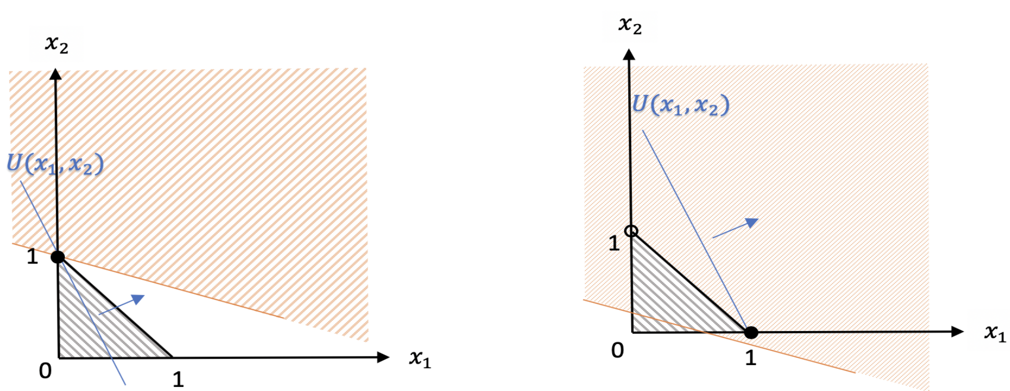
Figure W7: Dual-goal Programming Solution (Case 3)



Corresponding ε-constraint formulations.

If we use R as an additional constraint (Figure W8), we can now have either of the two solutions depending on the value of θ. Smaller values of θ lead to one of the two potential solutions point as optimal as shown in Figure W8(a). Larger values of θ lead to both points being feasible, but $(x_1, x_2) = (1, 0)$ is optimal as shown in Figure W8(b). Here, the ε-constraint formulation leads to a unique solution given θ.

Figure W8: Additional constraint $R(x_1, x_2) \leq \theta_R$ (Case 3)

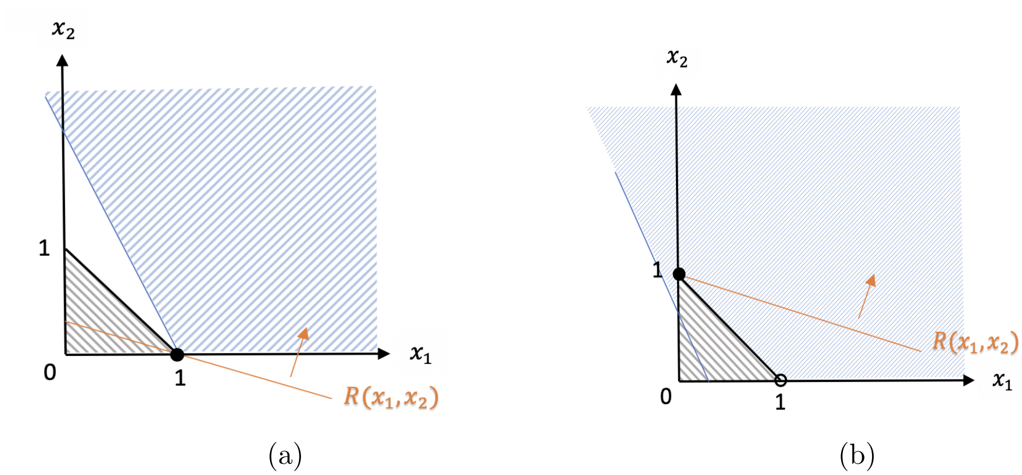


(a) (b)

Figure W9 illustrates the alternative ε-constraint formulation, wherein the value of θ_U

plays the same role.

Figure W9: Additional constraint $-U(x_1, x_2) \leq \theta_U$ (Case 3)



B.4 Case 4

Our final case illustrates the possibility of three possible solutions that are Pareto optimal. We show that the value of θ leads to different optimal solutions, highlighting the importance of the observed consideration set data.

Original problem.

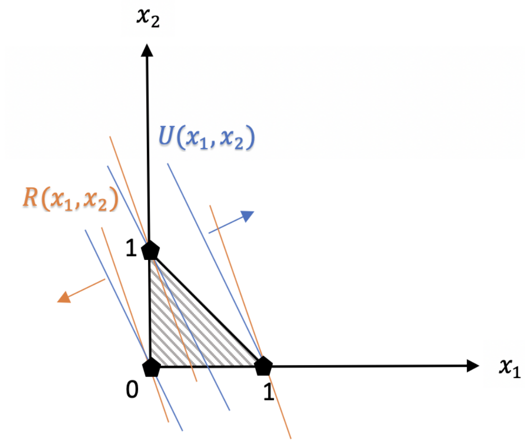
$$\max \ U(x_1, x_2) = 2x_1 + x_2$$

$$\min \ R(x_1, x_2) = 3x_1 + x_2$$

$$s.t. \quad x_1 + x_2 \leq 1, \quad x_1, x_2 \in \{0, 1\}$$

The three possible solutions are $(x_1^*, x_2^*) = (1, 0)$, $(x_1^*, x_2^*) = (0, 1)$, and $(x_1^*, x_2^*) = (0, 0)$, and no solution set dominates the others. Therefore all of them are Pareto optimal (Figure W10).

Figure W10: Dual-goal Programming Solution (Case 4)



Corresponding ε-constraint formulations.

As shown above, the optimal choice is dependent on the threshold value θ. Figure W11 illustrates the ε-constraint formulation with R given as an additional constraint, and Figure W12 illustrates the formulation when U serves as an additional constraint.

Figure W11: Additional constraint $R(x_1, x_2) \leq \theta_R$ (Case 4)

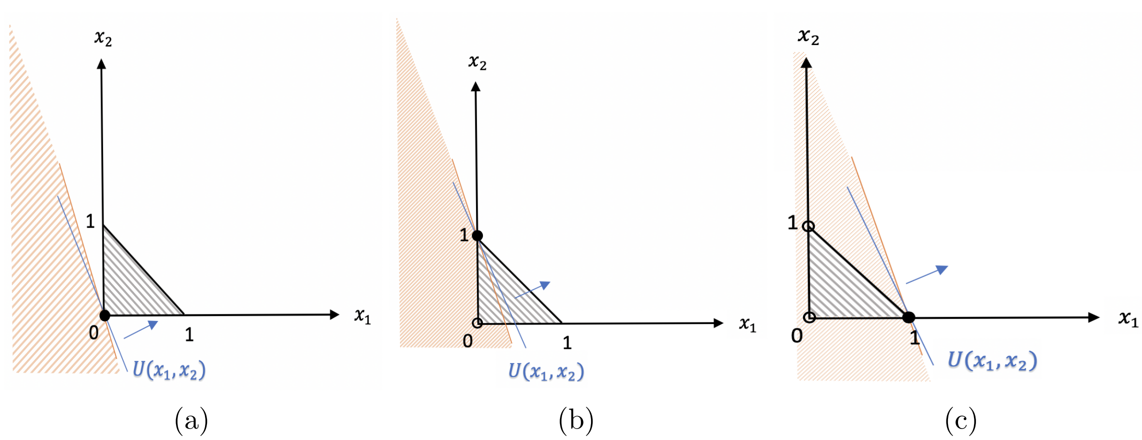
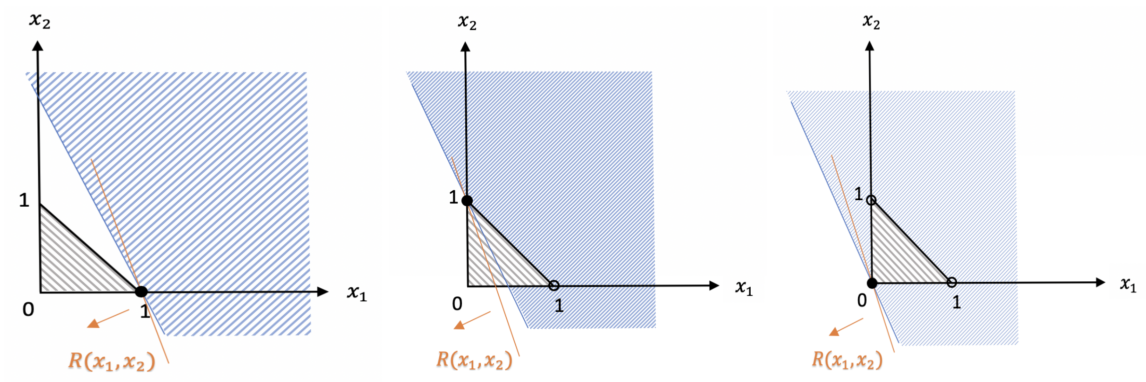


Figure W12: Additional constraint $-U(x_1, x_2) \leq \theta_U$ (Case 4)



(a) (b) (c)

The ε-constraint formulation always leads to a unique optimal solution given θ, even when there exists a set of Pareto optimal solution points in dual-goal optimization problem.

Web Appendix C: GENERALIZATION OF THE ε-CONSTRAINT FRAMEWORK

The model of dual-goal pursuit can be extended to multiple goal optimization problems. If there are N objective functions $\{f_n\}$ subject to $g_m(\boldsymbol{x}) \leq C_m(m = 1, \cdots, M)$ constraints, the multiple objective optimization problem is specified as:

$$\begin{aligned} &\min_{\boldsymbol{x}} \ f_1(\boldsymbol{x}) \quad \cdots \quad \min \ f_N(\boldsymbol{x}) \\ &s.t. \quad g_m(\boldsymbol{x}) \leq C_m, \ m = 1, \cdots, M \end{aligned}$$

We can reformulate the problem into optimizing one of the objective functions with the remaining $N-1$ goals put as additional constraints such that:

$$\begin{aligned} &\min_{\boldsymbol{x}} \ f_1(\boldsymbol{x}) \\ &s.t. \quad f_n(\boldsymbol{x}) \leq \varepsilon_n, \ n = 2, \cdots, N, \quad g_m(\boldsymbol{x}) \leq C_m, \ m = 1, \cdots, M \end{aligned}$$

Alternatively, we can optimize f_n with other objectives converted into additional constraints. The data required for estimation do not change as the number of goals increases, and the consideration set data is rationalized in terms of the conjunction (i.e., minimum subset) of the constraints in each formulation.

Web Appendix D: CLARK'S APPROXIMATION ALGORITHM

According to Clark (1961),

$$\begin{aligned}\mathbb{E}\Big[\max\{X_1, X_2, \cdots, X_m, \cdots, X_{M-1}X_M\}\Big] &= \mathbb{E}\Big[\max\{X_1, \cdots, X_m, \cdots, \max(X_{M-1}, X_M)\}\Big] \\ &\approx \mathbb{E}\Big[\max\{X_1, \cdots, X_m, \cdots, X_{M-2}, Y_{M-1}\}\Big]\end{aligned}$$

where Y_{M-1} is a two-moment Normal approximation to $\max(X_{M-1}, X_M)$.

If $X_{M-1} \sim N(\mu_{M-1}, \sigma^2_{M-1})$, $X_M \sim N(\mu_M, \sigma^2_M)$, the first moment of $\max(X_{M-1}, X_M)$ is

$$\nu_1 = \mu_{M-1}\Phi\left(\frac{\mu_{M-1}-\mu_M}{\sqrt{\sigma^2_{M-1}+\sigma^2_M}}\right) + \mu_M\Phi\left(-\frac{\mu_{M-1}-\mu_M}{\sqrt{\sigma^2_{M-1}+\sigma^2_M}}\right) + \left(\sqrt{\sigma^2_{M-1}+\sigma^2_M}\right)\phi\left(\frac{\mu_{M-1}-\mu_M}{\sqrt{\sigma^2_{M-1}+\sigma^2_M}}\right)$$

and the second moment of $\max(X_{M-1}, X_M)$ is

$$\begin{aligned}\nu_2 = (\mu^2_{M-1}+\sigma^2_{M-1})\Phi\left(\frac{\mu_{M-1}-\mu_M}{\sqrt{\sigma^2_{M-1}+\sigma^2_M}}\right) + (\mu^2_M+\sigma^2_M)\Phi\left(-\frac{\mu_{M-1}-\mu_M}{\sqrt{\sigma^2_{M-1}+\sigma^2_M}}\right) \\ + (\mu_{M-1}+\mu_M)\left(\sqrt{\sigma^2_{M-1}+\sigma^2_M}\right)\phi\left(\frac{\mu_{M-1}-\mu_M}{\sqrt{\sigma^2_{M-1}+\sigma^2_M}}\right)\end{aligned}$$

The expectation and variance of $\max(X_{M-1}, X_M)$ is

$$\mathbb{E}\Big[\max(X_{M-1}, X_M)\Big] = \nu_1$$

and

$$Var\Big[\max(X_{M-1}, X_M)\Big] = \nu_2 - \nu_1^2$$

Now we have $Y_{M-1} \overset{\text{approx}}{\sim} N(\nu_1, \nu_2 - \nu_1^2)$. The algorithm is generalized in the following.

Algorithm 1: Clark's approximation algorithm

Initialize ; $m = M$

while $m \geq 2$ **do**

If $X_{m-1} \sim N(\mu_{m-1}, \sigma^2_{m-1})$, $X_m \sim N(\mu_m, \sigma^2_m)$,

Calculate

$$\nu_1 = \mu_{m-1}\Phi\Big(\frac{\mu_{m-1}-\mu_m}{\sqrt{\sigma^2_{m-1}+\sigma^2_m}}\Big) + \mu_m\Phi\Big(-\frac{\mu_{m-1}-\mu_m}{\sqrt{\sigma^2_{m-1}+\sigma^2_m}}\Big) + \Big(\sqrt{\sigma^2_{m-1}+\sigma^2_m}\Big)\phi\Big(\frac{\mu_{m-1}-\mu_m}{\sqrt{\sigma^2_{m-1}+\sigma^2_m}}\Big)$$

$$\nu_2 = (\mu^2_{m-1}+\sigma^2_{m-1})\Phi\Big(\frac{\mu_{m-1}-\mu_m}{\sqrt{\sigma^2_{m-1}+\sigma^2_m}}\Big) + (\mu^2_m+\sigma^2_m)\Phi\Big(-\frac{\mu_{m-1}-\mu_m}{\sqrt{\sigma^2_{m-1}+\sigma^2_m}}\Big) + (\mu_{m-1}+\mu_m)\Big(\sqrt{\sigma^2_{m-1}+\sigma^2_m}\Big)\phi\Big(\frac{\mu_{m-1}-\mu_m}{\sqrt{\sigma^2_{m-1}+\sigma^2_m}}\Big)$$

Define Y_{m-1}

$$Y_{m-1} = \max(X_{m-1}, X_m) \overset{\text{approx}}{\sim} N(\nu_1, \nu_2 - \nu_1^2)$$

Update X_{m-1}

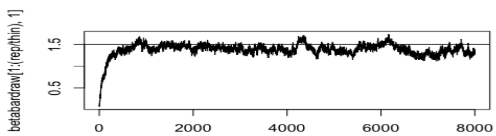
$$X_{m-1} = Y_{m-1}$$

$m = m - 1$

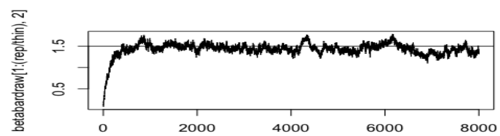
Output: $\mathbb{E}(X_m)$

Web Appendix E: SIMULATION RESULTS

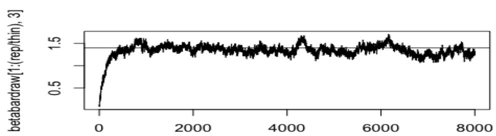
Figure W1: Simulation Results (D=4000)



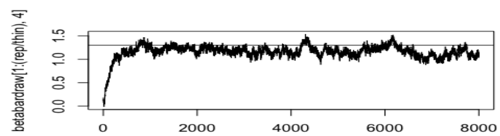
(a) β_1



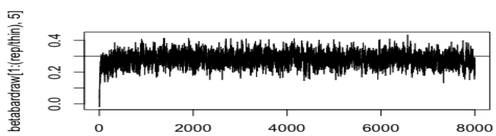
(b) β_2



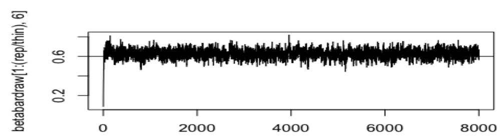
(c) β_3



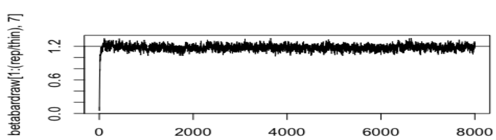
(d) β_4



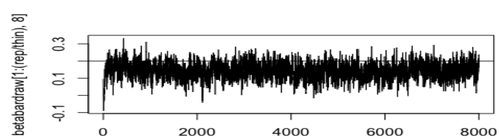
(e) β_5



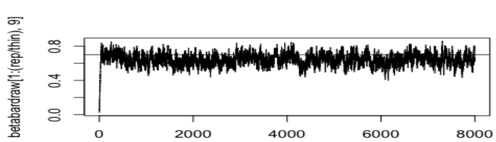
(f) β_6



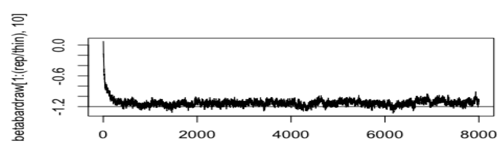
(g) β_7



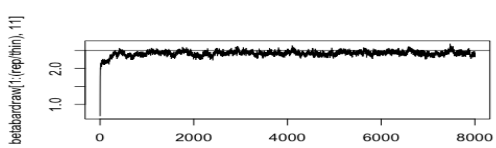
(h) β_8



(i) β_9



(j) β_{10}



(k) θ

Web Appendix F: ESTIMATION ALGORITHM

F.1 Notation

Respondents: $i = 1, \cdots, N$

Alternatives: $j = j_1, \cdots, j_J$

Attributes: $k = 1, \cdots, K$

Choice occasions: $t = 1, \cdots, T$

F.2 Estimation Procedure

Step 1. Initialize values for all variables to be inferred: $\boldsymbol{\beta}_i, \overline{\boldsymbol{\beta}}, \Sigma_{\boldsymbol{\beta}}$
where $\boldsymbol{\beta_i} = \begin{pmatrix} \beta_{i1}, & \cdots & \beta_{iK}, & \theta_i \end{pmatrix}^T$. We assume

$$\boldsymbol{\beta_i} \sim MVN\left(\overline{\boldsymbol{\beta}}, \quad \boldsymbol{\Sigma_\beta}\right)$$

Step 2. Draw $\eta_{ij}^d \overset{i.i.d}{\sim} EV(0,1)$ (D=4000)

Step 3. Generate $\boldsymbol{\beta_i}$ for $i = 1, 2, ..., N$ given $\overline{\boldsymbol{\beta}}$, and $\boldsymbol{\Sigma_\beta}$ via the random-walk Metropolis-Hastings algorithm:

(a) Draw candidate $\boldsymbol{\beta_i}^{\text{new}} \sim MVN(\boldsymbol{\beta_i}^{\text{old}}, s^2\boldsymbol{I})$. $\boldsymbol{\beta_i}^{\text{old}}$ is the previous value of $\boldsymbol{\beta_i}$ and s is the step size.

(b) Calculate $R_{ijt}^d \ \forall j, \forall t$

$$R_{ijt}^d = \mathbb{E}\Big[\max\big(\sum_{k=1}^{K} \mathbf{x}_{1tk}^{'}\boldsymbol{\beta}_{ik} + \eta_{i1}^d + \xi_{i1}, \cdots, \sum_{k=1}^{K} \mathbf{x}_{Jtk}^{'}\boldsymbol{\beta}_{ik} + \eta_{iJ}^d + \xi_{iJ}\big)\Big] - \sum_{k=1}^{K} \mathbf{x}_{jtk}^{'}\boldsymbol{\beta}_k - \eta_{ij}^d$$

using Clark's approximation. Here, R_{ijt}^d is a D-dimensional vector.

(c) For $\{j_1, \cdots, j_m \in C_{it}\}$, $\{j_{m+1}, \cdots, j_J \notin C_{it}\}$, evaluate approximate likelihood using Monte Carlo method:

$$\mathscr{L}_i = \prod_{t=1}^{T} \Bigg[\frac{1}{D} \sum_{d=1}^{D} \mathbb{I}(R^d_{ij_1t} \leq \theta_{i\kappa_1}, \cdots, R^d_{ij_mt} \leq \theta_{i\kappa_m}, \cdots, R^d_{i,j_{m+1}t} > \theta_{i\kappa_{m+1}}, \cdots, R^d_{ij_Jt} > \theta_{i\kappa_J})$$

$$\times \frac{1}{D'} \sum_{d'=1}^{D'} \mathbb{I}\Big\{ V_{ijt} + \eta^{d'}_{ijt} > V_{ij't} + \eta^{d'}_{ij't}, \quad \eta^{d'}_{ij_m} \in \big\{\boldsymbol{\eta} \,|\, R_{ijt} \leq \theta_{ij} \quad \forall j \in C\big\} \Big\} \Bigg]$$

where $\{\eta^{d'}_{ij}\}$ are realized errors such that $\mathbb{I}\{R^d_{ij_1t} \leq \theta_{i\kappa_1}, \cdots, R^d_{ij_mt} \leq \theta_{i\kappa_m}, \cdots, R^d_{i,j_{m+1}t} > \theta_{i\kappa_{m+1}}, \cdots, R^d_{ij_Jt} > \theta_{i\kappa_J}\} = 1$ and $D' = |\{\eta^{d'}_{ij}\}|$.

(d) Accept $\boldsymbol{\beta_i}^{\text{new}}$ with following probability:

$$\alpha = \min\left[1, \frac{\mathscr{L}_i\ (\boldsymbol{\beta_i}^{\text{new}}) \cdot \pi(\boldsymbol{\beta_i}^{\text{new}}|\overline{\boldsymbol{\beta}}, \Sigma_{\boldsymbol{\beta}})}{\mathscr{L}_i\ (\boldsymbol{\beta_i}^{\text{old}}) \cdot \pi(\boldsymbol{\beta_i}^{\text{old}}|\overline{\boldsymbol{\beta}}, \Sigma_{\boldsymbol{\beta}})}\right],$$

where $\pi(\cdot|\overline{\boldsymbol{\beta}}, \Sigma_{\boldsymbol{\beta}})$ is the density of the normal distribution with mean $\overline{\boldsymbol{\beta}}$ and variance $\Sigma_{\boldsymbol{\beta}}$.

(e) Generate $\overline{\boldsymbol{\beta}}$ and $\Sigma_{\boldsymbol{\beta}}$ given $\{\boldsymbol{\beta_i}\}$ via the following Bayesian multivariate regression (Rossi et al. 2005):

$$\{\boldsymbol{\beta_i}\} = \overline{\boldsymbol{\beta}} + \zeta_i, \quad \zeta_i \sim MVN(0, \Sigma_{\boldsymbol{\beta}}).$$

Prior distributions are given by $\overline{\boldsymbol{\beta}}|\Sigma_{\boldsymbol{\beta}} \sim MVN\big(\overline{\overline{\boldsymbol{\beta}}},\ \Sigma_{\boldsymbol{\beta}} \otimes A^{-1}\big)$ and $\Sigma_{\boldsymbol{\beta}} \sim \text{IW}(\nu, \nu \cdot \mathbf{I}_{\text{nvar}})$, where nvar denotes the dimension of $\overline{\boldsymbol{\beta}}$, and $\mathbf{I}_{\text{nvar}}$ is an (nvar×nvar) identity matrix.

(f) Repeat (a)-(e)

Web Appendix G: CALCULATION OF FIT MEASURES

G.1 Log-Marginal Density (LMD-NR)

Following Newton and Raftery (1994)'s,

$$LMD = \left[\frac{1}{R}\sum_{r=1}^{R}\frac{1}{l(\boldsymbol{B_r}|M)}\right]^{-1}$$

where R is the number of MCMC iterations, $\boldsymbol{B_r}$ is the r-th posterior draw of the parameters, M is the given model, and l is the likelihood values.

G.2 Log-Marginal Density (LMD-GD)

Following Gelfand and Dey (1994)'s,

$$LMD = \left[\frac{1}{R}\sum_{r=1}^{R}\frac{q(\boldsymbol{B_r})}{l(\boldsymbol{B_r}|M)p(\boldsymbol{B_r}|M)}\right]^{-1}$$

where q is an auxiliary density function (multivariate normal in our analysis).

G.3 Hit Probability

Consideration Set Hit Probability

$$= \frac{1}{N \times T \times R}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{r=1}^{R}\left[Pr\Big(\widehat{R_{ij_1tr}} < \widehat{\theta_{ij_1r}}, \cdots, \widehat{R_{ij_mtr}} < \widehat{\theta_{ij_mr}}, \widehat{R_{ij_{m+1}tr}} > \widehat{\theta_{ij_{m+1}r}} \cdots, \widehat{R_{ijtr}} > \widehat{\theta_{ij_Mr}}\Big)\right]$$

for $j_1, \cdots, j_m \in C$ and $j_{m+1}, \cdots, j_M \notin C$ where C is the consideration set.

Choice Hit Probability given Consideration Set

$$= \frac{1}{N \times T \times R}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{r=1}^{R}\prod_{j=1}^{J}\left[\frac{\exp(u_{ijtr}(\boldsymbol{B_r}|M))}{\sum_{j'\in C}\exp(u_{ij'tr}(\boldsymbol{B_r}|M))}\right]^{\mathbb{I}\{y_{it}=j\}}\left[1-\frac{\exp(u_{ijtr}(\boldsymbol{B_r}|M))}{\sum_{j'\in C}\exp(u_{ij'tr}(\boldsymbol{B_r}|M))}\right]^{\mathbb{I}\{y_{it}\neq j\}}$$

G.4 WAIC

Following Watanabe (2010)'s,

$$WAIC = -2\sum_{i=1}^{N} \log\left(\frac{1}{R}\sum_{r=1}^{R} p(y_i|\boldsymbol{B_r})\right) + 2\sum_{i=1}^{N} V_{r=1}^{R}\Big(\log p(y_i|\boldsymbol{B_r})\Big)$$

where V is the sample variance.