

Application of Nested Logit Model on Recommendation System

Wei-Yu Fan*, Yu-Chan Chen

May 2024

Abstract

我們探討了 nested logit model 在推薦系統上的應用。我們透過隨機效用模型與 nested logit distribution 將消費者點擊商品是為一隨機事件，藉由消費者對於不同商品的喜好程度而調整，在有限的廣告欄位下決定商品的組合並推播給消費者以最大化消費者的點籍機率。並且我們推導出此模型有一個 closed-form solution，並提供與之對應的演算法。

Keywords: Random Utility Model, Nested Logit Model,
Recommendation System

*Graduate student in Department of Economics, National Taiwan University.
Email address: entrencemania@gmail.com

1 Introduction

隨機效用模型一直以來都是推薦系統中的重要模型之一，透過隨機效用模型我們可以將消費者與商品的關係建模，並且透過此模型我們可以推導出消費者對於商品的偏好，而其中以 logit model 最廣為人知，並且衍伸出處理 3 個以上選擇的 multilogit model，logit family 的優勢在於其機率易於計算，並且能搭配回歸模型使用。McFadden (1978) 提出 generalized extreme value (GEV) distribution，也就是將 logit model 推廣到更一般的形式，而 nested logit model 是 GEV distribution 的一個特例，其特點在於可以處理商品間的相關性，推薦系統中此相關性假設尤為重要，設想在 4 個廣告欄位中分別投放了：螢幕、顯示卡、鍵盤、耳機四則廣告，此種投放策略能提升想要購買科技產品之消費者的點擊率，但是當下沒有科技產品購買意願的消費者就沒有充分的誘因點擊，此種投放策略就是沒有考慮到這四種商品之間的高度相關性。Nested logit model 就是建立在此種情況下的模型，其特點在於可以處理商品間的相關性，透過將商品分群 (clustered)，使得商品間的相關性更為明顯，進而提升推薦系統的效果。

2 Related Literature

3 Fundamental Concepts

3.1 Random Utility Model

離散選擇是計量經濟學描繪消費者行為的一個重要方法，其中隨機效用模型允許消費者在選擇商品時有一定的隨機性，這種隨機性可以解釋消費者在同一個選擇集中選擇不同商品的原因，假設代表性消費者對於兩商品 $i \in \{1, 2\}$ 的效用分別為

$$\begin{aligned} U_1 &= V_1 + \epsilon_1, \\ U_2 &= V_2 + \epsilon_2, \quad (\epsilon_1, \epsilon_2)^T \sim ((0, 0)^T, \Sigma). \end{aligned}$$

其中 U_i 代表消費者對於商品 i 的效用， V_i 代表商品 i 的固定效用， (ϵ_i, ϵ_j) 服從一期望值為 0 之聯合分布，將消費者的選擇以 dependent variable y 表示，則消費者選擇商品 1 的機率為

$$\begin{aligned} P(y = 1) &= P(U_1 > U_2) = P(\epsilon_1 - \epsilon_2 > V_2 - V_1) \\ &= \int_{\epsilon_1 - \epsilon_2 > V_2 - V_1} f(\epsilon_1, \epsilon_2) d\epsilon_1 d\epsilon_2. \end{aligned}$$

3.2 Logit Model

當我們假設 $(\epsilon_1, \epsilon_2) \stackrel{\text{iid}}{\sim}$ type-I Generalized Extreme Value Distribution，則此模型會退化為 logit model，消費者選擇商品 1 的機率為

$$P(y = 1) = \frac{e^{V_1}}{e^{V_1} + e^{V_2}}.$$

當我們假設 $V_2 = 0$ 時也就是將 $i=2$ 解釋為不購買商品，則此模型會退化為 binary logit model，消費者選擇商品 1 的機率為

$$P(y = 1) = \frac{e^{V_1}}{e^{V_1} + 1}.$$

此外若將商品數量推廣為 n 種， $(\epsilon_1, \epsilon_2, \dots, \epsilon_n) \stackrel{\text{iid}}{\sim}$ type-I Generalized Extreme Value Distribution 則此模型會退化為 multinomial logit model，消費者選擇商品 i 的機率為

$$\begin{aligned} P(y = i) &= P(U_i \geq U_k), \quad \forall k \in \{1, 2, \dots, n\} \\ &= \frac{e^{V_i}}{\sum_{k=1}^n e^{V_k}}. \end{aligned}$$

3.3 Generalized Extreme Value Model

McFadden (1978) 提出 generalized extreme value (GEV) distribution，也就是將 logit model 推廣到更一般的形式，此分布的優勢在於應用在隨機效用模型時其選擇機率易於計算，只要隨機效用 $(\epsilon_1, \epsilon_2, \dots, \epsilon_n)$ 之聯合累積分布

函數符合以下性質

$$F(\epsilon_1, \epsilon_2, \dots, \epsilon_n) = \exp[-G(e^{-\epsilon_1}, e^{-\epsilon_2}, \dots, e^{-\epsilon_n})], \quad (1)$$

其中滿足 aggregate function $G(y_1, y_2, \dots, y_n)$ 的充要條件為

1. Nonnegativity
2. Homogeneity of degree 1
3. Mixed partial derivatives are continuous and nonpositive for even order and nonnegativity for odd order
4. $G(y_1, y_2, \dots, y_n) \rightarrow \infty$ when $y_i \rightarrow \infty$

則此模型的選擇機率為

$$P(y = i) = e^{V_i} \frac{G_i(e^{-V_1}, e^{-V_2}, \dots, e^{-V_n})}{G(e^{-V_1}, e^{-V_2}, \dots, e^{-V_n})}, \quad (2)$$

$$G_i(y_1, y_2, \dots, y_n) := \frac{\partial G(y_1, y_2, \dots, y_n)}{\partial y_i}.$$

當選定 $G(y_1, y_2, \dots, y_n) = \sum_{j=1}^n y_j$ 為 aggregate function 時，此模型會退化為 multinomial logit model

$$\begin{aligned} P(y = i) &= e^{V_i} \frac{G_i(e^{-V_1}, e^{-V_2}, \dots, e^{-V_n})}{G(e^{-V_1}, e^{-V_2}, \dots, e^{-V_n})} \\ &= \frac{e^{V_i}}{\sum_{k=1}^n e^{V_k}}. \end{aligned}$$

3.4 Nested Logit Model

傳統的 logit model 假設不同商品間的隨機效用是獨立的，但在真實世界中這是一個不合理的假設，因此我們允許同一種類的商品間有相關性但不同種類的商品間依然互相獨立，e.g. 同樣是科技產品的螢幕、顯示卡、鍵盤、耳機的隨機效用具有相關性，但是科技產品的螢幕與食品的巧克力是互相獨立的。

假設所有的商品被分為 M 個類別，第 m 群的商品數量為 N_m ，在不失一般性的情況下，假設 $N_1 = N_2 = \dots = N_M = N$ ，此時消費者的選擇 $y \in \mathbf{R}^2$ (i.e. $y = (i, j)$ 為消費者選擇第 i 類別中的第 j 項商品)。使用 GEV model 來建構符合此場景的模型就是 nested logit model，其定義與性質如下：

Definition 3.1 (Nested logit model). Under random utility model, if the joint cumulative distribution function of $(\epsilon_{1,1}, \epsilon_{1,2}, \dots, \epsilon_{M,N})$ satisfies

$$\begin{aligned} F(\epsilon_{1,1}, \epsilon_{1,2}, \dots, \epsilon_{M,N}) &= \exp[-G(e^{-\epsilon_{1,1}}, e^{-\epsilon_{1,2}}, \dots, e^{-\epsilon_{M,N}})] \\ &= \exp\left\{-\sum_{m=1}^M \left[\sum_{n=1}^N \exp\left(-\frac{\epsilon_{m,n}}{\tau_m}\right)\right]^{\tau_m}\right\}, \end{aligned}$$

this model is called nested logit model.

The parameters τ_m are called dissimilarity parameters which are required to satisfy $0 \leq \tau_m \leq 1$ for all $m = 1, 2, \dots, M$.

Proposition 3.1 (Probability of nested logit model). Under nested logit

model, the probability of consumer choosing item (i,j) is

$$P(y = (i, j)) = \frac{e^{\frac{V_{i,j}}{\tau_i}}}{\sum_{n=1}^N e^{\frac{V_{i,n}}{\tau_i}}} \frac{(\sum_{n=1}^N e^{\frac{V_{i,n}}{\tau_i}})^{\tau_i}}{\sum_{m=1}^M (\sum_{n=1}^N e^{\frac{V_{m,n}}{\tau_m}})^{\tau_m}},$$

this proposition can be proved by equation 2.

Proposition 3.2 (Evaluation of probability). Directly from proposition 3.1,

$$\begin{aligned} P(y = (i, j)) &= \frac{e^{\frac{V_{i,j}}{\tau_i}}}{\sum_{n=1}^N e^{\frac{V_{i,n}}{\tau_i}}} \frac{(\sum_{n=1}^N e^{\frac{V_{i,n}}{\tau_i}})^{\tau_i}}{\sum_{m=1}^M (\sum_{n=1}^N e^{\frac{V_{m,n}}{\tau_m}})^{\tau_m}} \\ &= P(y = (i, j) \mid y = (i, *)) P(y = (i, *)), \end{aligned}$$

the first fraction is the conditional probability of choosing item (i,j) given that the consumer chooses category i, and the second fraction is the probability of choosing category i.

Proposition 3.3 (Covariance and correlation of nested logit model). Under nested logit model, the covariance and correlation of random utility of **different** items (i,j) and (k,l) are

$$Cov(\epsilon_{i,j}, \epsilon_{k,l}) = \begin{cases} \frac{\pi^2}{6}(1 - \tau_i^2), & \text{if } i=k \\ 0, & \text{otherwise} \end{cases}$$

$$Cor(\epsilon_{i,j}, \epsilon_{k,l}) = \begin{cases} 1 - \tau_i^2, & \text{if } i=k \\ 0, & \text{otherwise} \end{cases}$$

4 Methodologies

Suppose there are M nests each have N alternatives, the benefit of nested logit model is that the decision is decomposed into two steps: first, the consumer chooses a nest, then the consumer chooses an alternative within the chosen nest. In the scenario of recommendation system, the consumer clicks on an ad instead of buying a product. Moreover, the consumer can choose not to click on any ad, which can also be considered as a kind of ad. The "not-click" ad is a special case which is the only choice in the not-click nest. Let 0'th nest be the not-click nest, the utility of consumer click on any of these ads is

$$U_{i,j} = \begin{cases} V_{i,j} + \epsilon_{i,j}, & \text{click } j\text{'th ad of } i\text{'th nest} \\ V_0 + \epsilon_0, & \text{not click any ad.} \end{cases} \quad (3)$$

Notice that we only observe the choice of consumer, but not the utility of consumer. Which means equation 3 cannot be identified. A common way to solve this problem is to shift the whole model by subtracting V_0 (i.e. let $\tilde{U}_{i,j} = U_{i,j} - V_0$). For readability, we will omit the tilde in the following discussion. Instead of model 3 we discuss the following model:

$$U_{i,j} = \begin{cases} V_{i,j} + \epsilon_{i,j}, & \text{click } j\text{'th ad of } i\text{'th nest} \\ \epsilon_0, & \text{not click any ad.} \end{cases} \quad (4)$$

Another issue is indentification of the dissimilarity parameter τ_i . Heiss(2002) finds out if a nest contains only one alternative, the dissimilarity parameter of this nest cannot be indentified. In this case, only not-click nest is affected. The solution is rather simple, we set $\tau_0 = 1$.

Combining the discussion above, the probability of consumer click on ad (i,j) is

$$P(y = (i, j)) = \frac{e^{\frac{V_{i,j}}{\tau_i}}}{\sum_{n=1}^N e^{\frac{V_{i,n}}{\tau_i}}} \frac{(\sum_{n=1}^N e^{\frac{V_{i,n}}{\tau_i}})^{\tau_i}}{1 + \sum_{m=1}^M (\sum_{n=1}^N e^{\frac{V_{m,n}}{\tau_m}})^{\tau_m}}, \quad (5)$$

and the probability of consumer not click on any ad is

$$P(y = (0, 0)) = \frac{1}{1 + \sum_{m=1}^M (\sum_{n=1}^N e^{\frac{V_{m,n}}{\tau_m}})^{\tau_m}}. \quad (6)$$

The third issue is that the consumer only views limited number of ads, which means the consumer can only choose from a subset of all ads. Therefore, we define the set of ads that the consumer can choose from as an exposure matrix $\mathbf{S} \in \mathbf{R}^{M \times N}$.

Definition 4.1 (Exposure Matrix). $\mathbf{S} \in \mathbf{R}^{M \times N}$ is the exposure matrix if and

only if

$$\mathbf{s}_{i,j} = \begin{cases} 1, & \text{if ad (i,j) is exposed to consumer} \\ 0, & \text{otherwise.} \end{cases}$$

Futhermore, let $\mathbf{s}_i \in \mathbf{R}^N$ denote the exposure vector that the consumer can choose from nest i . The relation between \mathbf{S} and \mathbf{s}_i is that $\mathbf{s}_i = \mathbf{S}^T \mathbf{e}_i$, where $\mathbf{e}_i \in \mathbf{R}^M$ is a vector with 1 at i -th position and 0 elsewhere.

In the following sections, we will discuss several cases from simple to complex. The first case is that the consumer as an all-knowing being, which means the consumer can observe any number of ads at ease. The second case is that the consumer is a normal being, and the exposure constaint is that the consumer can obeserve limited number of ads per nest. The third case is about the substitute effect within ads. We will discuss the effect of another one ad being exposed to the consumer and the impact on the probability of consumer click on the original ad. The fourth case is about the more familiar case, the consumer can only observe limited number of ads in total.

4.1 Case 1: All-knowing Consumer

The goal is to maximize the probability of consumer click on any ad. In other words, we aim to minimize the probability of consumer not click on any ad. The optimization problem is

$$\begin{aligned}
\text{Objective funtion: } \quad & \min_{\mathbf{s}} P(y = (0, 0)) \\
&= \frac{1}{1 + \sum_{m=1}^M (\sum_{n=1}^N \mathbf{s}_{\mathbf{m}, \mathbf{n}} e^{\frac{V_{m,n}}{\tau_m}})^{\tau_m}} \quad (7)
\end{aligned}$$

To achieve notation simplicity, we define $\mathbf{H} \in \mathbf{R}^{M \times N}$ as the exponent utility matrix.

Definition 4.2 (Exponent Utility Matrix). $\mathbf{H} \in \mathbf{R}^{M \times N}$ is the exponent utility matrix if and only if

$$\mathbf{H}_{\mathbf{i}, \mathbf{j}} = e^{\frac{V_{i,j}}{\tau_i}}.$$

Similarly, we define $\mathbf{h}_{\mathbf{i}} \in \mathbf{R}^N$ as the exponent utility vector of nest \mathbf{i} . The relation between \mathbf{H} and $\mathbf{h}_{\mathbf{i}}$ is that $\mathbf{h}_{\mathbf{i}} = \mathbf{H}^T \mathbf{e}_{\mathbf{i}}$. Therefore we can rewrite the optimization problem as

$$\begin{aligned}
\text{Objective funtion: } \quad & \min_{\mathbf{s}} P(y = (0, 0)) \\
&= \frac{1}{1 + \sum_{m=1}^M (\mathbf{h}_{\mathbf{m}}^T \mathbf{s}_{\mathbf{m}})^{\tau_m}} \quad (8)
\end{aligned}$$

The optimal solution \mathbf{S}^* is

$$\begin{aligned}
\mathbf{S}^* &= \arg \min_{\mathbf{s}} P(y = (0, 0)) = \arg \min_{\mathbf{s}} \frac{1}{1 + \sum_{m=1}^M (\mathbf{h}_{\mathbf{m}}^T \mathbf{s}_{\mathbf{m}})^{\tau_m}} \\
&= \arg \max_{\mathbf{s}} \sum_{m=1}^M (\mathbf{h}_{\mathbf{m}}^T \mathbf{s}_{\mathbf{m}})^{\tau_m}. \quad (9)
\end{aligned}$$

Since the components of exponent utility vector \mathbf{h}_m are all positive combining with the fact that the components of exposure vector \mathbf{s}_m are all binary, the optimal solution should be that the all-knowing consumer exposed to all ads to achieve the maximum probability of consumer click on any ad

$$\mathbf{s}_{i,j}^* = 1, \quad \forall i \in \{1, 2, \dots, M\}, j \in \{1, 2, \dots, N\}. \quad (10)$$

4.2 Case 2: Limited Exposure in each Nest

Case 2 is case 1 with the exposure constraint. The problem can be written as

$$\begin{aligned} \text{Objective function:} \quad & \min_{\mathbf{s}} P(y = (0, 0)) = \frac{1}{1 + \sum_{m=1}^M (\mathbf{h}_m^T \mathbf{s}_m)^{\tau_m}}, \\ \text{Subject to:} \quad & \sum_{n=1}^N \mathbf{s}_{m,n} \leq K, \quad \forall m \in \{1, 2, \dots, M\}, \end{aligned} \quad (11)$$

where K is some integer less than N .

Case 1 implies that the inner solution (i.e. $\sum_{n=1}^N \mathbf{s}_{m,n} < K$) never be the optimal. Hence we focus on the bounded solution specticular. The optimal solution is rather simple as well. Without loss of generality, we can reorder the components of \mathbf{H} such that $\mathbf{H}_{m,1} \geq \mathbf{H}_{m,2} \geq \dots \geq \mathbf{H}_{m,N}$, $\forall m \in \{1, 2, \dots, M\}$. The optimal solution is

$$\mathbf{s}_{m,n}^* = \begin{cases} 1, & \text{if } n \leq K \\ 0, & \text{otherwise.} \end{cases} \quad (12)$$

The interpretation is that the consumer should be exposed to the K highest exponent utility ads in each nest.

4.3 Case 3: Substitute Effect

Here we discuss the effect of another one ad being exposed to the consumer and the impact on the probability of consumer click on the original ad. We define the substitute effect of ad (o,p) on ad (i,j) as

Definition 4.3 (Substitute Effect).

$$\text{SE}[(o, p) \rightarrow (i, j)] = P(y = (i, j) \mid \mathbf{S}_{o,p} = 1) - P(y = (i, j) \mid \mathbf{S}_{o,p} = 0).$$

The substitute effect can be calculated by the following formula

$$P(y = (i, j) \mid \mathbf{S}_{o,p} = 0) = \frac{\mathbf{H}_{i,j}}{\mathbf{h}_i^T \mathbf{s}_i} \frac{(\mathbf{h}_i^T \mathbf{s}_i)^{\tau_i}}{1 + \sum_{m=1}^M (\mathbf{h}_m^T \mathbf{s}_m)^{\tau_m}}$$

$$P(y = (i, j) \mid \mathbf{S}_{o,p} = 1) = \begin{cases} \frac{\mathbf{H}_{i,j}}{\mathbf{h}_i^T \mathbf{s}_i} \frac{(\mathbf{h}_i^T \mathbf{s}_i)^{\tau_i}}{1 + \sum_{m=1, m \neq o}^M (\mathbf{h}_m^T \mathbf{s}_m)^{\tau_m} + (\mathbf{h}_o^T \mathbf{s}_o + \mathbf{H}_{o,p})^{\tau_o}}, & o \neq i, \\ \frac{\mathbf{H}_{i,j}}{\mathbf{h}_i^T \mathbf{s}_i + \mathbf{H}_{o,p}} \frac{(\mathbf{h}_i^T \mathbf{s}_i + \mathbf{H}_{o,p})^{\tau_i}}{1 + \sum_{m=1, m \neq o}^M (\mathbf{h}_m^T \mathbf{s}_m)^{\tau_m} + (\mathbf{h}_o^T \mathbf{s}_o + \mathbf{H}_{o,p})^{\tau_o}}, & o = i. \end{cases}$$

Assume $\mathbf{H}_{o,p}$ is rather small, the first order Taylor expansion

of $P(y = (i, j) \mid \mathbf{S}_{o,p} = 1)$ at $\mathbf{H}_{o,p} = 0$ is

$$P(y = (i, j) \mid \mathbf{S}_{o,p} = 1) \approx \begin{cases} P(y = (i, j) \mid \mathbf{S}_{o,p} = 0) \left[1 - \frac{\mathbf{H}_{o,p}}{\mathbf{h}_o^T \mathbf{s}_o} \frac{\tau_o (\mathbf{h}_o^T \mathbf{s}_o)^{\tau_o}}{1 + \sum_{m=1}^M (\mathbf{h}_m^T \mathbf{s}_m)^{\tau_m}} \right], & o \neq i, \\ P(y = (i, j) \mid \mathbf{S}_{o,p} = 0) \left[1 - \frac{\mathbf{H}_{o,p}}{\mathbf{h}_o^T \mathbf{s}_o} \left(\frac{\tau_o (\mathbf{h}_o^T \mathbf{s}_o)^{\tau_o}}{1 + \sum_{m=1}^M (\mathbf{h}_m^T \mathbf{s}_m)^{\tau_m}} + (1 - \tau_o) \right) \right], & o = i. \end{cases} \quad (13)$$

From equation 13, we can derive the substitute effect of ad (o,p) on ad (i,j) as

$$\begin{aligned} \text{SE}[(o, p) \rightarrow (i, j)] &= P(y = (i, j) \mid \mathbf{S}_{o,p} = 1) - P(y = (i, j) \mid \mathbf{S}_{o,p} = 0) \\ &\approx \begin{cases} -P(y = (i, j) \mid \mathbf{S}_{o,p} = 0) \frac{\mathbf{H}_{o,p}}{\mathbf{h}_o^T \mathbf{s}_o} \frac{\tau_o (\mathbf{h}_o^T \mathbf{s}_o)^{\tau_o}}{1 + \sum_{m=1}^M (\mathbf{h}_m^T \mathbf{s}_m)^{\tau_m}}, & o \neq i, \\ -P(y = (i, j) \mid \mathbf{S}_{o,p} = 0) \frac{\mathbf{H}_{o,p}}{\mathbf{h}_o^T \mathbf{s}_o} \left[\frac{\tau_o (\mathbf{h}_o^T \mathbf{s}_o)^{\tau_o}}{1 + \sum_{m=1}^M (\mathbf{h}_m^T \mathbf{s}_m)^{\tau_m}} + (1 - \tau_o) \right], & o = i. \end{cases} \end{aligned} \quad (14)$$

The result is actually intuitive. First, the substitute effect of ad (o,p) on ad (i,j) is negative in both cases, which means that the probability of consumer click on ad (i,j) decreases when another ad is exposed to the consumer. This makes sense because in the nested logit model, all the ads are substitutes to each other with respect to probability. Second, the magnitude of substitute effect is larger($1 - \tau_o$ term in equation 14) when the old ad shares the same nest with the new ad. This is reasonable because the ads in the same nest are more similar to each other, which means the consumer is more likely to substitute the new ad with the old one. For example, the substitute effect of a new iPhone ad on a Samsung mobile ad is larger than

the substitute effect of a new iPhone ad on a cereal ad due to the fact that iPhone and Samsung mobile are both in the smartphone category but cereal isn't.

4.4 Case 4: Limited Exposure in Total

Case 3 is case 1 with a different exposure constraint. The problem can be written as

$$\begin{aligned}
\text{Objective function: } \min_{\mathbf{s}} P(y = (0, 0)) &= \frac{1}{1 + \sum_{m=1}^M (\mathbf{h}_m^T \mathbf{s}_m)^{\tau_m}}, \\
\text{Subject to: } \sum_{m=1}^M \sum_{n=1}^N \mathbf{S}_{m,n} &\leq K, \quad \text{where } K \text{ is some integer less than } M \cdot N.
\end{aligned} \tag{15}$$

The constraint can be written as $\sum_{m=1}^M \mathbf{s}_m^T \mathbf{s}_m \leq K$, which seems like the nostalgia L2 regularization in machine learning. The difference is that the L2 regularization is a continuous constraint while the constraint in random utility problem is a discrete constraint. This kind of problem is called the knapsack problem in combinatorial optimization and it is NP-hard in general. Nonetheless, in our setting the optimal solution \mathbf{S}^* is P-hard.

In the first step, sort components of each row of \mathbf{H} in descending order i.e. $\mathbf{H}_{m,1} \geq \mathbf{H}_{m,2} \geq \dots \geq \mathbf{H}_{m,N}$, $\forall m \in \{1, 2, \dots, M\}$. To proceed further, we need the following lemma.

Lemma 4.1 (Monotonicity of Choice). \mathbf{S}^* is the optimal solution. If $\mathbf{S}_{i,j}^* = 1$,

then

$$\mathbf{S}_{i,j-1}^* = 1$$

Proof. The proof is by contradiction. Suppose $\mathbf{S}_{i,j-1}^* = 0$, then we can construct a new solution \mathbf{S}' by setting $\mathbf{S}'_{i,j} = 0$ and $\mathbf{S}'_{i,j-1} = 1$. Since the objective function is the same as case 1, we can rewrite the objective function as

$$\text{Alternative Objective function: } \max_{\mathbf{S}} \sum_{m=1}^M (\mathbf{h}_m^T \mathbf{s}_m)^{\tau_m}.$$

The difference between alternative value functions achieved by \mathbf{S}' and \mathbf{S}^* is

$$\begin{aligned} & \sum_{m=1}^M (\mathbf{h}_m^T \mathbf{s}'_m)^{\tau_m} - \sum_{m=1}^M (\mathbf{h}_m^T \mathbf{s}_m^*)^{\tau_m} \\ &= (\mathbf{h}_i^T \mathbf{s}'_i)^{\tau_i} - (\mathbf{h}_i^T \mathbf{s}_i^*)^{\tau_i} \\ &= \left(\sum_{n \neq j, n \neq j-1}^N \mathbf{H}_{i,n} \mathbf{S}'_{i,n} + \mathbf{H}_{i,j-1} \right)^{\tau_i} - \left(\sum_{n \neq j, n \neq j-1}^N \mathbf{H}_{i,n} \mathbf{S}_{i,n}^* + \mathbf{H}_{i,j} \right)^{\tau_i} \\ &= \left(\sum_{n \neq j, n \neq j-1}^N \mathbf{H}_{i,n} \mathbf{S}_{i,n}^* + \mathbf{H}_{i,j-1} \right)^{\tau_i} - \left(\sum_{n \neq j, n \neq j-1}^N \mathbf{H}_{i,n} \mathbf{S}_{i,n}^* + \mathbf{H}_{i,j} \right)^{\tau_i} > 0. \end{aligned}$$

Therefore, the alternative solution \mathbf{S}' is better than the optimal solution \mathbf{S}^* , which contradicts the assumption that \mathbf{S}^* is the optimal solution. \square

Lemma 4.2 (Bounded Constraint). Optimal solution \mathbf{S}^* always satisfies the bounded constraint i.e.

$$\sum_{m=1}^M \sum_{n=1}^N \mathbf{S}_{m,n}^* = K.$$

Proof. This is proof by contradiction. Suppose $\sum_{m=1}^M \sum_{n=1}^N \mathbf{S}_{m,n}^* < K$, then

we can construct a new solution \mathbf{S}' by setting $\mathbf{S}'_{ij} = 1$ where $\mathbf{S}^*_{ij} = 0$. Using the alternative objective function in lemma 4.1, we can show that the alternative solution \mathbf{S}' is better than the optimal solution \mathbf{S}^* , which contradicts the assumption that \mathbf{S}^* is the optimal solution. \square

5 Conclusion

References

- [1] Irandoust, Manuchehr. (2019). *House Prices and Unemployment: An Empirical Analysis of Causality*, International Journal of Housing Markets and Analysis. 12. 148-164.
- [2] Shiou-Yen Chu (2018). *Macroeconomic policies and housing market in Taiwan*, International Review of Economics & Finance, Volume 58, 404-421.