Nested Logit Derivatives for NLS Gradient

C. Luke Watson*

April 9, 2024

1 Small Model

Consider a model of demand where various agents, t, make choices over products, j, that are sold by various brands, f. Let there be four inside goods and one outside good. Let there be three nests $\{N_0, N_1, N_2\}$, such that we have the following nesting structure: $\{(0), (1,2), (3,4)\}$. Let d_{jt} be the unconditional share of the good, let $d_{jt|n}$ be the nest-conditional share of the good, where $n = N(j)$ is implied for each good j, and let d_n be the nest share. Let demand be calculated as $D_{jt} = d_{jt} \cdot LS_t$, where LS_t is a scaling factor that translates shares to demand-quantities. Let V_{jt} be the utility value of a good, where:

$$V_{jt} = \begin{cases} Y_{jt}\beta^1 + X_j\beta^2 + Z_t\beta^3 + X_jZ_t\beta^4 + F_j\beta^5, & j > 0 \\ W_t\lambda^1 + Z_t\lambda^2 + W_tZ_t\lambda^3, & j = 0, \end{cases} \tag{1}$$

where Y is agent-product characteristics, X are product characteristics, F is a brand FE, Z are 'inside' agent characteristics, and W are 'outside' agent characteristics.

The above along with the nested-logit assumptions yields the following unconditional product share:

$$d_{1t} = \frac{\left(\mathrm{e}^{V_{1t}/\rho_1}\right)\left(\mathrm{e}^{V_{1t}/\rho_1} + \mathrm{e}^{V_{2t}/\rho_1}\right)^{\rho_1 - 1}}{\left(\mathrm{e}^{V_{0t}}\right) + \left(\mathrm{e}^{V_{1t}/\rho_1} + \mathrm{e}^{V_{2t}/\rho_1}\right)^{\rho_1} + \left(\mathrm{e}^{V_{3t}/\rho_2} + \mathrm{e}^{V_{3t}/\rho_2}\right)^{\rho_2}} \tag{2}$$

*Watson: Center for Financial Research, Federal Deposit Insurance Corporation, email: chwatson@fdic.gov. The views expressed here do not represent the views of the FDIC or US Government.

2 Gradient

We wish to minimize the squared loss function:

$$m(\beta, \lambda, \rho) = \frac{1}{2 \cdot N_b} \sum_{b \in N_b} \left(\ln \left(D_b(\beta, \lambda, \rho) \right) - \ln \left(D_b^{\text{Data}} \right) \right)^2 . \tag{3}$$

Let $\theta = (\beta, \lambda, \rho)$ with K elements, then the gradient is a vector $g = \nabla_\theta m$, such that:

$$\begin{aligned} g_k = \frac{\partial m(\theta)}{\partial \theta^k} &= \frac{1}{N_b} \sum_j \left(\frac{\partial \ln(D_j(\theta))}{\partial \theta^k} \right) = \frac{1}{N_b} \sum_j \left(\frac{\partial \ln(\sum_{t \in \mathcal{S}_j} LS_t d_{jt}(\theta))}{\partial \theta^k} \right) \\ &= \frac{1}{N_b} \sum_j \left(\sum_{t \in \mathcal{S}_j} \left\{ \left(\frac{D_{jt}}{D_j} \right) \left(\frac{\partial d_{jt} / \partial \theta^k}{d_{jt}} \right) \right\} \right) . \end{aligned} \tag{4}$$

3 Partial Derivatives wrt the Parameters

Here, we display the derivatives of d_{1t} necessary for the gradient.

$$\frac{\partial d_{1t}}{\partial \beta^k} = \frac{d_{1t}}{\rho_1} \left[\frac{\partial V_{1t}}{\partial \beta^k} + (\rho_1 - 1) \left(\sum_{j \in N_1} \frac{\partial V_{jt}}{\partial \beta^k} d_{jt|N_1} \right) - \rho_1 \left(\sum_n \sum_{j \in n} \frac{\partial V_{jt}}{\partial \beta^k} d_{jt} \right) \right] \tag{5}$$

$$= \frac{d_{1t}}{\rho_1} \left[\Gamma_{1t}^{1,\beta^k} + \Gamma_{N_1,t}^{2,\beta^k} + \Gamma_t^{3,\beta^k} \right] \tag{6}$$

$$\frac{\partial d_{1t}}{\partial \lambda^k} = \frac{\partial V_{0t}}{\partial \lambda^k} d_{0t} d_{1t} \tag{7}$$

$$= d_{1t} \left[\Gamma_t^{1,\lambda^k} \right] \tag{8}$$

$$\frac{\partial d_{1t}}{\partial \rho_1} = -\frac{d_{1t}}{\rho_1^2} \left[V_{1t} + (\rho_1(1 - d_{N_1,t}) - 1) \left(\sum_{j \in N_1} V_{jt} d_{jt|N_1} \right) - (\rho_1^2 (1 - d_{N_1,t})) \cdot \Lambda_{N_1,t} \right] \tag{9}$$

$$= -\frac{d_{1t}}{\rho_1^2} \left[\Gamma_{1t}^{1,\rho_1} + \Gamma_{N_1,t}^{2,\rho_1} \right] \tag{10}$$

$$\frac{\partial d_{1t}}{\partial \rho_2} = \frac{d_{1t}}{\rho_2^2} \left[\rho_2 \left(\sum_{j \in N_2} V_{jt} d_{jt} \right) - \rho_2^2 (d_{N_2,t}) \cdot \Lambda_{N_2,t} \right] \tag{11}$$

$$= \frac{d_{1t}}{\rho_2^2} \left[\Gamma_{N_2,t}^{1,\rho_2} \right] , \tag{12}$$

where $\Lambda_{n,t} = \ln \left(\sum_{j \in n} \mathrm{e}^{V_{jt}/\rho_n} \right)$, and the $\{\Gamma\}$ terms are implicitly defined.