

# Appendix: Choosing an $n$ -Pack of Substitutable Products

## A Proofs

### THEOREM 2

*Proof. Part I: Optimality of stopping condition.* Consider an  $n$ -pack with quantities  $k_i$  for alternative  $i$  that satisfies the stopping rule (9). Then for any other pack with quantities  $k'_i$ , we can define two sets:  $I^+ = \{i : k_i > k'_i\}$  and  $I^- = \{j : k_j < k'_j\}$ . We have  $V(k_1, k_2, \dots, k_M) = \ln(n!) - \sum_{i=1}^M \ln(k_i!) + \sum_{i=1}^M k_i U_i$  and  $V(k'_1, k'_2, \dots, k'_M) = \ln(n!) - \sum_{i=1}^M \ln(k'_i!) + \sum_{i=1}^M k'_i U_i$ . Now observe that

$$\begin{aligned} V(k_1, k_2, \dots, k_M) - V(k'_1, k'_2, \dots, k'_M) &= \sum_{i \in I^+} \ln(k'_i!) - \ln(k_i!) + (k_i - k'_i)U_i \quad (\text{A.1}) \\ &+ \sum_{j \in I^-} \ln(k'_j!) - \ln(k_j!) + (k_j - k'_j)U_j. \end{aligned}$$

Furthermore, observe that

$$\sum_{i \in I^+} (k_i - k'_i) = - \sum_{j \in I^-} (k_j - k'_j) \quad (\text{A.2})$$

because any gain in units across  $I^+$  must be exactly matched by losses across  $I^-$  to maintain an  $n$ -pack with exactly  $n$  units.

Suppose now we (1) split the  $U'_i$ s and  $U'_j$ s in (A.1) into individual terms having unit coefficients (i.e.,  $(k_i - k'_i)U_i = \sum_{l=1}^{k_i - k'_i} U_i$  and  $(k_j - k'_j)U_j = -\sum_{l=1}^{k'_j - k_j} U_j$ ) and (2) cancel like terms in the log-factorials. The term  $\ln(k'_i!) - \ln(k_i!) + (k_i - k'_i)U_i$  in the first summation on the right hand side of (A.1) can thus be expressed as

$$\begin{aligned} \ln(k'_i!) - \ln(k_i!) + (k_i - k'_i)U_i &= (U_i - \ln(k_i)) + (U_i - \ln(k_i - 1)) + \dots \quad (\text{A.3}) \\ &+ (U_i - \ln(k'_i + 1)) \end{aligned}$$

whereas the term  $\ln(k'_j!) - \ln(k_j!) + (k_j - k'_j)U_j$  in the second summation on the right hand

side of (A.1) can be expressed as

$$\begin{aligned} \ln(k'_j!) - \ln(k_j!) + (k_j - k'_j)U_j &= (\ln(k_j + 1) - U_j) + (\ln(k_j + 2) - U_j) + \dots \\ &+ (\ln(k'_j) - U_j). \end{aligned} \quad (\text{A.4})$$

If we do this for each term in each summation, we will obtain a total of  $\sum_{i \in I^+} (k_i - k'_i)$  terms of each form (see (A.2)). We can then pair each term of the form  $U_i - \ln(k_i - l_i)$  for  $i \in I^+$ ,  $l_i \in \mathbb{N}^0$ ,  $0 \leq l_i \leq k_i - k'_i - 1$  from (A.3) with a term of the form  $-U_j + \ln(k_j + m_j)$  for  $j \in I^-$ ,  $m_j \in \mathbb{N}$ ,  $1 \leq m_j \leq k'_j - k_j$  from (A.4). These pairings can be done in any manner. We can thus rewrite (A.1) using a single summation where each (paired) term has the general form

$$\begin{aligned} U_i - U_j - \ln(k_i - l_i) + \ln(k_j + m_j) & \quad (\text{A.5}) \\ i \in I^+, l_i \in \mathbb{N}^0, l_i \leq k_i - k'_i - 1 & \\ j \in I^-, m_j \in \mathbb{N}, 1 \leq m_j \leq k'_j - k_j & \end{aligned}$$

But by the stopping condition (9), with the roles of  $i$  and  $j$  interchanged, we must have

$$\begin{aligned} U_i - \ln(k_i) - U_j + \ln(k_j + 1) &\geq 0 \text{ all } i, j \\ \implies U_i - \ln(k_i - l_i) - U_j + \ln(k_j + m_j) &\geq 0 \text{ all } i, j \end{aligned}$$

which implies all the terms in (A.5) are non-negative and thus so is the summation in (A.1). Thus  $V(k_1, k_2, \dots, k_M) - V(k'_1, k'_2, \dots, k'_M) \geq 0$  as was desired.

*Part II: Convergence in at most  $n$  swaps.* To show the algorithm converges in at most  $n$  swaps, we will show that an alternative that gains a unit will never lose a unit, and an alternative that loses a unit will never gain back a unit (note this implies each unit can be moved at most once). To do this, we first replace each utility  $U_i$  with the quantity  $U_i - i \cdot \varepsilon$  where  $\varepsilon$  is a non-Archimedean infinitesimal, a positive number that is smaller than any number in the base field. This creates a strict ordering in the utilities used in (6,7) so that ties are broken by the non-Archimedean term. This is equivalent to breaking ties in (6) by selecting the alternative with the *smallest* index, and breaking ties in (7) by selecting the alternative with the *largest* index.

We first show that an alternative that gains a marginal unit can never lose that marginal unit. Suppose  $g$  is a maximizer to (6) and therefore a “gainer” in the current swap and  $l$  is the minimizer to (7) and therefore a “loser” in the current swap. If  $g = l$ , then it is easy to see the optimality conditions are met and we are done. We therefore assume  $g \neq l$ . We

must have

$$U_g - g \cdot \varepsilon - \ln(k_g + 1) > U_i - i \cdot \varepsilon - \ln(k_i + 1) \text{ all } i \neq g. \quad (\text{A.6})$$

$$U_l - l \cdot \varepsilon - \ln(k_l) < U_j - j \cdot \varepsilon - \ln(k_j) \text{ } j \neq l. \quad (\text{A.7})$$

The gain ( $\Delta$ ) in our objective function for the current swap is

$$\Delta = U_g - g \cdot \varepsilon - \ln(k_g + 1) - U_l + l \cdot \varepsilon + \ln(k_l) > 0. \quad (\text{A.8})$$

The alternative  $g$  cannot be the minimizer for (7)—and hence the loser—in the *next* iteration of the algorithm unless the optimality conditions have been achieved. For if  $g$  is the minimizer of (7) in the next swap, then using  $k_g + 1$  units for alternative  $g$  and  $k_l - 1$  units for alternative  $l$ , we may apply (5) to calculate the possible gains in the objective function at the next iteration as

$$U_i - i \cdot \varepsilon - \ln(k_i + 1) - (U_g - g \cdot \varepsilon - \ln(k_g + 1)) \quad i \neq l, g \quad (\text{A.9})$$

$$U_l - l \cdot \varepsilon - \ln(k_l) - (U_g - g \cdot \varepsilon - \ln(k_g + 1)). \quad (\text{A.10})$$

The potential gains in (A.9) are all negative by (A.6); the potential gain in (A.10) is negative by (A.8). This means the optimality conditions have been met (and we are done), or else  $g$  is not the minimizer for (7) in the next iteration of our algorithm. If it is not the minimizer, then there are two possibilities: case (i)  $g$ , and only  $g$ , gains additional units in all future swaps; case (ii) some other alternative, say  $g'$  ( $g' \neq g$ ), gains one or more units at some point.

For case (i), there is nothing to prove because this is consistent with our premise ( $g$  never loses the  $(k_g + 1)^{\text{st}}$  unit). For case (ii), we know by (A.6) that  $U_g - g \cdot \varepsilon - \ln(k_g + 1) > U_{g'} - g' \cdot \varepsilon - \ln(k_{g'} + 1)$ , and so  $U_g - g \cdot \varepsilon - \ln(k_g + 1) > U_{g'} - g' \cdot \varepsilon - \ln(k_{g'} + m)$  for any  $m \in \{1, 2, \dots\}$ . Thus  $g$  could never be a minimizer in (7) with  $(k_g + 1)$  units of inventory once it has been a maximizer in (6). Since  $g$  and  $k_g$  were arbitrary, this proves that any marginal unit gained is never lost.

We now show that an alternative that loses a marginal unit can never gain back that marginal unit. We claim the alternative  $l$  cannot be a maximizer for (A.6)—and hence a gainer—in the *next* iteration of the algorithm unless the optimality conditions have been achieved. For if  $l$  is a maximizer in the next swap, then using  $k_g + 1$  units for alternative

$g$  and  $k_l - 1$  units for alternative  $l$ , we may apply (5) to calculate the possible gains in the objective function at the next iteration as

$$U_l - l \cdot \varepsilon - \ln(k_l) - (U_j - j \cdot \varepsilon - \ln(k_j)) \quad j \neq l, g \quad (\text{A.11})$$

$$U_l - l \cdot \varepsilon - \ln(k_l) - (U_g - g \cdot \varepsilon - \ln(k_g + 1)) \quad (\text{A.12})$$

The potential gains in (A.11) are all negative by (A.7); the potential gain in (A.12) is negative by (A.8). This means the optimality conditions have been met (and we are done), or else  $l$  is not the maximizer for (A.6) in the next iteration of our algorithm. If  $l$  is not the maximizer, there are two remaining possibilities: case (i)  $l$ , and only  $l$ , loses additional units in all future swaps; case (ii) some other alternative, say  $l'$  ( $l' \neq l$ ), loses one or more units at some point.

For case (i), there is nothing to prove because this is consistent with our premise ( $l$  never gains back the  $l^{\text{th}}$  unit). For case (ii), we know by (A.7) that  $U_l - l \cdot \varepsilon - \ln(k_l) < U_{l'} - l' \cdot \varepsilon - \ln(k_{l'})$  and so  $U_l - l \cdot \varepsilon - \ln(k_l) < U_{l'} - l' \cdot \varepsilon - \ln(k_{l'} - m)$  for any  $m \in \{1, 2, \dots\}$ ,  $k_{l'} - m \geq 0$ . Thus  $l$  could never be a maximizer in (A.6) with  $(k_l - 1)$  units of  $l$  once it has been a minimizer in (A.7). Since  $l$  and  $k_l$  were arbitrary, this proves that any marginal unit lost is never gained back.

Because an alternative that gains a unit can never lose that unit, and an alternative that loses a unit can never gain back that unit, each unit in the initial feasible solution can be moved at most once. This means the algorithm converges in at most  $n$  swaps.

□

### THEOREM 3

*Proof.* We first create a “dummy alternative,” say  $U_{M+1}$ , which is much less attractive than any alternative in the current list. For example, one could set  $U_{M+1} = -\Theta$ , where  $\Theta$  is larger than any number in the base field. Now apply the algorithm of Theorem 2 with the initial (dummy) solution  $k_i = 0$  for  $i = 1, 2, \dots, M$  and  $k_{M+1} = n$ . (This starting solution is analogous to that used by the “big M” method in LP’s simplex algorithm.) Alternative  $M+1$  will always be the minimum in (7), which means units will be removed from this alternative one at a time until there are no units left. The first unit removed will *always* go to alternative 1 because it solves (6). Observe that this must be the optimal 1-pack; for if the dummy

alternative initially had  $n=1$  units, the algorithm would terminate. The second unit removed always goes to the alternative that solves (6) with the values  $k_1 = 1, k_i = 0$  for  $i = 2, \dots, M$ . Observe that the alternative receiving the second unit is independent of  $n$  ( $n \geq 2$ ) and the resulting 2-pack must be the optimal 2-pack; for if the dummy alternative initially had  $n=2$  units, the algorithm would terminate. Proceeding in this fashion, at iteration  $n + 1$ , we compute the optimal  $(n + 1)$ -pack by adding a unit to the alternative that maximizes  $U_i - \ln(k_i^* + 1)$ , where the values  $k_i^*$  are the quantities already determined for the optimal  $(n)$ -pack at the previous iteration. Observe that  $n$  only determines how many swaps are made, and the sequence of swaps is otherwise independent of  $n$ .  $\square$

## THEOREM 4

*Proof.* To prove (i), assume the optimal solution satisfies  $k_i = 0$  and  $k_j > 0$  with  $U_i > U_j$ . Then swapping the two values increases the linear utility term in (1) without affecting the choice premium. This violates the optimality of the assumed solution and means the optimal solution must be a contiguous set of the consumer's favorite alternatives.

To prove (ii), assume for some optimal  $n$ -pack  $(k_1^*, k_2^*, \dots, k_M^*)$  that  $k_i^* < k_j^*$  for some  $U_i > U_j$ . Then one could swap the quantities as in (i) to increase the value function (4). (Observe the linear utility term would increase while the choice premium remained unchanged.) this violates the optimality assumption.

To prove (iii), consider the optimality conditions in (9). For any two alternatives  $i$  and  $j$  we therefore have  $U_i - \ln(k_i^* + 1) \leq U_j - \ln(k_j^*)$ , which can be rearranged to yield the result.  $\square$

## THEOREM 5

*Proof.* For  $t = 1$ , the index set  $I_1(1, k_1, k_2, \dots, k_M)$  reduces to a set of  $m + 1$  vectors, each having dimension  $(M + 1)$ . One vector is  $(1, 0, 0, \dots)$ , which captures the selection of the outside option; the remaining  $m$  vectors have a "1" in the position of the alternative included in the  $n$ -pack and 0 elsewhere. The value formula for  $t = 1$  reduces to the well-known expected value formula for the alternative having maximum utility, which is (see , for instance,

Ghulam Ali, 2008)

$$V_1(k_1, k_2, \dots, k_M) = \ln \left[ \sum_{(x_0, x_1, \dots, x_M) \in I_1(1, k_1, k_2, \dots, k_M)} e^{\sum_{j=0}^M x_j U_j} \right] + \gamma. \quad (\text{A.13})$$

(Here,  $\gamma$  is the Euler-Mascheroni constant,  $\gamma \approx .577$ .) To prove the general case, suppose it is true for  $t$ . Then for  $t + 1$ , the optimal policy is to choose the alternative that maximizes current utility plus utility-to-go, i.e., the maximum of  $U_j + \epsilon_j + V_t(k_1, k_2, k_j - 1, \dots, k_M)$  for those  $j$  having  $k_j > 0$  and also  $U_0 + \epsilon_0 + V_t(k_1, k_2, \dots, k_M)$ . This implies we must have (using (A.13) with  $U_j$  replaced by  $U_j + V_t(k_1, k_2, k_j - 1, \dots, k_M)$  and  $U_0$  replaced by  $U_0 + V_t(k_1, k_2, \dots, k_M)$ ):

$$V_{t+1}(k_1, k_2, \dots, k_M) = \ln \left[ \left( \sum_{\{j: k_j > 0\}} e^{U_j + V_t(k_1, k_2, k_j - 1, \dots, k_M)} \right) + e^{U_0 + V_t(k_1, k_2, \dots, k_M)} \right] + \gamma, \quad (\text{A.14})$$

which by the induction step

$$= \ln \left[ \left\{ \sum_{\{j: k_j > 0\}} \sum_{(x_0, x_1, \dots, x_M) \in I_t(t, k_1, k_2, \dots, k_j - 1, \dots, k_M)} \frac{t!}{x_0! x_1! \dots x_M!} e^{U_j + \sum_{l=0}^M x_l U_l} \right. \right. \\ \left. \left. + \sum_{(x_0, x_1, \dots, x_M) \in I_t(t, k_1, k_2, \dots, k_M)} \frac{t!}{x_0! x_1! \dots x_M!} e^{U_0 + \sum_{l=0}^M x_l U_l} \right\} \cdot e^{t\gamma} \right] + \gamma \quad (\text{A.15})$$

Grouping like terms, observe that the coefficient of the general term  $e^{a_0 U_0 + a_1 U_1 + \dots + a_M U_M}$  ( $0 \leq a_i \leq k_i$ ,  $a_i \in \mathbb{N}^0$ ,  $\sum_{j=0}^M a_j = t + 1$ ) is made up of contributions from the summations in (A.15). To capture these contributions, define the function

$$\delta(z) = \begin{cases} 0 & \text{if } z = 0 \\ 1 & \text{if } z > 0. \end{cases}$$

Then the total contribution from the summations in (A.15) to the coefficient of  $e^{a_0 U_0 + a_1 U_1 + \dots + a_M U_M}$  can be calculated as follows:

$$\begin{aligned}
& \frac{\delta(a_0) \cdot t!}{\max(a_0 - 1, 0)! a_1! \dots a_M!} + \frac{\delta(a_1) \cdot t!}{a_0! \max(a_1 - 1, 0)! \dots a_M!} + \\
& + \frac{\delta(a_2) \cdot t!}{a_0! a_1! \max(a_2 - 1, 0)! \dots a_M!} + \dots + \frac{\delta(a_M) \cdot t!}{a_0! a_1! a_2! \dots \max(a_M - 1, 0)!} \\
= & \frac{t! (a_0 + a_1 + \dots + a_M)}{a_0! a_1! a_2! \dots a_M!} \\
= & \frac{(t + 1)!}{a_0! a_1! a_2! \dots a_M!},
\end{aligned}$$

which is precisely the coefficient needed to make the formula (13) correct for  $V_{t+1}(k_1, k_2, \dots, k_M)$ .

A nearly identical proof works for the canonical case where  $k_i \in \mathbb{N}^+$ ,  $\sum_{i=1}^M k_i = n$ , and the number of consumption occasions is  $n$  (although one can also prove this result quite easily by induction on the size,  $n$ , of the  $n$ -pack). In the previous proof, one eliminates  $j = 0$ ,  $U_0$ , and  $x_0$  from the analysis. The index set reduces to

$$I_t(k_1, k_2, \dots, k_M) = \left\{ (x_1, x_2, \dots, x_M) : \sum_{i=1}^M x_i = t, 0 \leq x_i \leq k_i, x_i \in \mathbb{N}^0 \ i = 1, 2, \dots, M \right\}$$

for  $t \leq n$  and the outside option terms in equations A.14 and A.15 disappear. The index set for the value function in period  $n$  is  $I_n(k_1, k_2, \dots, k_M)$ , which contains a single vector,  $(k_1, k_2, \dots, k_M)$ . The optimal value function reduces to  $V_n(k_1, k_2, \dots, k_M) = \ln \left( \frac{n!}{k_1! k_2! \dots k_M!} e^{\sum_{i=1}^M k_i U_i} \right) + n\gamma$ , which is the formula stated in Theorem 1. The optimal policy stated for the canonical model follows from inserting this value function in the general form of the optimal policy:  $\max_{j: k_j > 0} (U_j + \epsilon_j + V_{n-1}(k_1, k_2, k_j - 1, \dots, k_M))$ . The latter is algebraically equivalent to  $\max_{j: k_j > 0} (V_n(k_1, k_2, k_j, \dots, k_M) - \ln(n) + \ln(k_j) + \epsilon_j) = \max_{j: k_j > 0} (\ln(k_j) + \epsilon_j)$ .  $\square$

## THEOREM 6

*Proof.* We recall that  $k_1^* \geq k_2^* \geq \dots \geq k_M^*$  must be true for the optimal  $n$ -pack of the canonical model. An identical permutation argument (the one described in Theorem 4) proves  $q_1^* \geq q_2^* \geq \dots \geq q_M^*$  must be true for the generalized model as well. To eliminate the use of subscripts whenever possible, let  $q^* = (q_1^*, q_2^*, \dots, q_M^*)$ , and let  $e_j \in \mathbb{R}^M$  be the

standard unit basis vectors, i.e.,  $e_1 = (1, 0, \dots, 0)$ ,  $e_2 = (0, 1, \dots, 0)$ , etc. We will work with the function  $\exp(V_t(q^*))$  because it eliminates the logarithm in (13) and allows us to work directly with the summation term. Observe that optimizing  $\exp(V_t(q))$  is equivalent to optimizing  $V_t(q)$ .

To prove the theorem, we argue by contradiction. Suppose there are indices  $i$  and  $j$  where  $q_j^* > k_j^*$  and  $q_i^* < k_i^*$  for  $j < i$ . We will now show that the  $n$ -pack with  $q_j^* - e_j + e_i$  in the generalized model has greater expected utility than the “optimal”  $n$ -pack  $q^*$ . To do so, observe that  $\exp(V_t(q^*))$  and  $\exp(V_t(q^* - e_j + e_i))$  share many terms in common that can be differenced out. Because  $t \geq n$ ,  $\exp(V_t(q^*))$  has  $\prod_{l=1}^M (q_l^* + 1)$  terms corresponding to the different of levels  $(0, 1, \dots, q_l^*)$  for each product alternative  $l$  (the outside option can be thought of as providing “padding” for the different combinations of products consumed from the  $n$ -pack). Note that  $\exp(V_t(q^* - e_j + e_i))$  has *more* terms than  $\exp(V_t(q^*))$  because  $q_j^* > q_i^*$  ( $q_j^* > k_j^* \geq k_i^* > q_i^*$ ). In fact, a careful accounting of terms reveals that  $\exp(V_t(q^* - e_j + e_i))$  and  $\exp(V_t(q^*))$  share  $\binom{q_j^*}{q_i^*} \prod_{l \neq j} (q_l^* + 1)$  identical terms (corresponding to identical consumption levels for all alternatives). However,  $\exp(V_t(q^*))$  has  $\prod_{l \neq j} (q_l^* + 1)$  unique terms (those corresponding to the fixed level  $x_j = q_j^*$  in equation (13)), whereas  $\exp(V_t(q^* - e_j + e_i))$  has  $\binom{q_j^*}{q_i^*} \prod_{l \neq j, i} (q_l^* + 1)$  unique terms (those corresponding to levels  $x_i = q_i^* + 1$  and  $x_j = 0, 1, \dots, q_j^* - 1$ ). Because  $q_j^* > k_j^* \geq k_i^* > q_i^*$ , we must have  $q_j^* \geq q_i^* + 2$  and so  $\binom{q_j^*}{q_i^*} \prod_{l \neq j, i} (q_l^* + 1) > \prod_{l \neq j} (q_l^* + 1)$ .

We now consider the difference  $\exp(V_t(q^*)) - \exp(V_t(q^* - e_j + e_i))$ , which eliminates the common terms. The unique terms for  $\exp(V_t(q^*))$  can subsequently be paired with a proper subset of those from  $\exp(V_t(q^* - e_j + e_i))$  in such a way that the number of times the outside option is used is identical between paired terms. Let  $x_l$  represent the level of alternative  $l$  (including  $l = 0$ ) for the terms in  $\exp(V_t(q^*))$ , and let  $x'_l$  represent the level of alternative  $l$  for the terms in  $\exp(V_t(q^* - e_j + e_i))$ . The matching we propose goes as follows:

$$\begin{aligned} x_l = x'_l \ l \neq i, j; \ x_j = q_j^*; \ x_i = q_i^* \text{ paired with } x'_l = x_l \ l \neq i, j; \ x'_j = q_j^* - 1; \ x'_i = q_i^* + 1 \\ x_l = x'_l \ l \neq i, j; \ x_j = q_j^*; \ x_i = q_i^* - 1 \text{ paired with } x'_l = x_l \ l \neq i, j; \ x'_j = q_j^* - 2; \ x'_i = q_i^* + 1 \\ x_l = x'_l \ l \neq i, j; \ x_j = q_j^*; \ x_i = q_i^* - 2 \text{ paired with } x'_l = x_l \ l \neq i, j; \ x'_j = q_j^* - 3; \ x'_i = q_i^* + 1 \\ \text{etc.} \end{aligned}$$

Note that the number of times the outside option is selected is the same within each pair because the number of products consumed from the  $n$ -pack is the same within each pair. Also, whereas all the terms for  $\exp(V_t(q^*))$  are paired off,  $\exp(V_t(q^* - e_j + e_i))$  still has additional positive terms that are unpaired. The general pair has the form  $x_l = x'_l \ l \neq i, j; \ x_j = q_j^*; \ x_i = q_i^* - z$  paired with  $x'_l = x_l \ l \neq i, j; \ x'_j = q_j^* - z - 1; \ x'_i = q_i^* + 1$  for  $z = 0, 1, \dots, q_i^*$ . Now analyzing the difference in values for the general pair, we observe it

equals

$$\begin{aligned}
& \frac{e^{\sum_{l \neq i, j} x_l U_l}}{\prod_{l \neq i, j} x_l!} \left[ \frac{e^{q_j^* U_j} e^{(q_i^* - z) U_i}}{q_j^*! (q_i^* - z)!} - \frac{e^{(q_j^* - z - 1) U_j} e^{(q_i^* + 1) U_i}}{(q_j^* - z - 1)! (q_i^* + 1)!} \right] \\
&= \frac{e^{\sum_{l \neq i, j} x_l U_l}}{\prod_{l \neq i, j} x_l!} \frac{e^{(q_i^* - z) U_i} e^{(q_j^* - z - 1) U_j}}{(q_i^* - z)! (q_j^* - z - 1)!} \left[ \frac{e^{(z+1) U_j}}{q_j^*!} - \frac{e^{(z+1) U_i}}{(q_i^* + 1)!} \right]. \tag{A.16}
\end{aligned}$$

We can now show the term in brackets is non-positive. To do so, we will make use of the optimality conditions of the canonical  $n$ -pack, which require

$$Max_j (U_j - \ln(k_j + 1)), j = 1, \dots, M \leq Min_i (U_i - \ln(k_i)), i = 1, \dots, M).$$

(This requires swapping  $i$  and  $j$  in equation (9)). Exponentiating the optimality condition means, in particular, that the canonical  $n$ -pack must satisfy

$$\frac{e^{U_j}}{k_j^* + 1} \leq \frac{e^{U_i}}{k_i^*},$$

which implies

$$\frac{e^{U_j}}{q_j^*} \leq \frac{e^{U_i}}{q_i^* + 1},$$

and thus

$$\begin{aligned}
\frac{e^{U_j}}{q_j^* - 1} &\leq \frac{e^{U_i}}{q_i^*} \\
\frac{e^{U_j}}{q_j^* - 2} &\leq \frac{e^{U_i}}{q_i^* - 1} \\
&\vdots
\end{aligned}$$

Multiplying the left hand sides of the first  $z + 1$  of these inequalities and then doing the same for the right hand sides implies

$$\frac{e^{(z+1) U_j}}{q_j^*!} \leq \frac{e^{(z+1) U_i}}{(q_i^* + 1)!}.$$

This implies the bracketed term in the difference equation (A.16) is always non positive. Because  $\exp(V_t(q^* - e_j + e_i))$  has additional positive terms, this implies  $\exp(V_t(q^*)) < \exp(V_t(q^* - e_j + e_i))$ , which proves the theorem. □

The following lemma is used in the proof of Theorem 7.

**Lemma 1.** Let  $a_i \geq 0, b_i \geq 0$  for  $i = 0, 1, \dots, m$ . Define the index sets  $S^+ = \{i : a_i > b_i\}$  and  $S^- = \{i : a_i \leq b_i\}$ . Then if  $\sum a_i \geq \sum b_i$ , we must have  $\sum_{i \in S^+} \beta_i a_i + \sum_{i \in S^-} \theta_i a_i \geq \sum_{i \in S^+} \beta_i b_i + \sum_{i \in S^-} \theta_i b_i$  for all  $0 \leq \theta_i \leq 1, \beta_i \geq 1$ .

*Proof.* By the conditions of the lemma, we must have  $\sum_{i \in S^+} (a_i - b_i) \geq \sum_{i \in S^-} -(a_i - b_i) \geq 0$ , and so  $\sum_{i \in S^+} \beta_i (a_i - b_i) \geq \sum_{i \in S^-} -\theta_i (a_i - b_i)$ , which is a re-arrangement of the stated result.  $\square$

## THEOREM 7

*Proof.* Without loss of generality, we take  $U_0 = 0$ . This means the additive constant appearing in  $V_t$  is changed from  $t\gamma$  to  $t(\gamma + U_0)$  (see equation (14)). We set  $\lambda = \gamma + U_0$  to simplify notation. Moreover, we choose to work with  $\exp[V(K)]$  and prove  $\exp[V_{t+1}(K) + V_{t-1}(K)] \leq \exp[V_t(K) + V_t(K)]$ . The proof is by induction on the size of the  $n$ -pack.

For  $n = 1$ , assume product  $j$  is the selected product. Then

$$\begin{aligned} \exp(-2t\lambda) \exp[V_{t+1}(K) + V_{t-1}(K)] &= (1 + (t+1)e^{U_j}) (1 + (t-1)e^{U_j}) \\ &= 1 + 2te^{U_j} + (t^2 - 1)e^{2U_j} \\ &\leq 1 + 2te^{U_j} + (t^2)e^{2U_j} \\ &= \exp(-2t\lambda) \exp[V_t(K) + V_t(K)] \end{aligned}$$

Assume it is true for all packs  $K$  of size  $n - 1$  or less. Observe that this means for any integers  $t, j$  with  $t \geq 1$  and  $0 \leq j \leq t - 1$

$$\frac{\exp[V_t(K)]}{\exp[V_{t-1}(K)]} \leq \frac{\exp[V_{t-j}(K)]}{\exp[V_{t-j-1}(K)]}.$$

Thus for  $t > t'$  and any integer  $k$  such that  $t - k \geq t' + k$ , we must also have

$$\begin{aligned} \exp(V_{t-k}(K) + V_{t'+k}(K)) &\geq \exp(V_{t-k+1}(K) + V_{t'+k-1}(K)) \\ &\geq \exp(V_{t-k+2}(K) + V_{t'+k-2}(K)) \\ &\vdots \\ &\geq \exp(V_t(K) + V_{t'}(K)) \end{aligned} \tag{A.17}$$

For any  $n$ -pack  $K = (k_1, k_2, \dots, k_M)$ , consider any product having a *positive* quantity.

Suppose  $j$  is one such product. Define  $\widehat{K}_j = (k_1, \dots, k_{j-1}, 0, k_{j+1}, \dots, k_M)$ . Observe that  $\widehat{K}_j$  must have  $n - 1$  items or less. Because  $U_0 = 0$ , we may write the value function (see (14)) as

$$\begin{aligned}
\exp[V_t(K)] &= \exp(t\lambda) \sum_{x_i \leq k_i, i \geq 1} \frac{t!}{(t - \sum_{i=1}^M x_i)!} \frac{\exp\left(\sum_{i=1}^M x_i U_i\right)}{\prod_{i=1}^M x_i!} \\
&= \exp(t\lambda) \sum_{z=0}^{k_j} \sum_{x_i \leq k_i, i \geq 1, i \neq j} \frac{t!}{(t - z - \sum_{i=1, i \neq j}^M x_i)!} \frac{\exp(zU_j) \exp\left(\sum_{i=1, i \neq j}^M x_i U_i\right)}{z! \prod_{i=1, i \neq j}^M x_i!} \\
&= \exp(t\lambda) \sum_{z=0}^{k_j} \frac{t! \exp(zU_j)}{z!(t-z)!} \sum_{x_i \leq k_i, i \geq 1, i \neq j} \frac{(t-z)!}{(t-z - \sum_{i=1, i \neq j}^M x_i)!} \frac{\exp\left(\sum_{i=1, i \neq j}^M x_i U_i\right)}{\prod_{i=1, i \neq j}^M x_i!} \\
&= \exp(t\lambda) \sum_{z=0}^{k_j} \frac{t! \exp(zU_j)}{z!(t-z)!} \exp\left(V_{t-z}(\widehat{K}_j)\right) \exp((z-t)\lambda) \\
&= \sum_{z=0}^{k_j} \binom{t}{z} \exp(z(U_j + \lambda)) \exp\left(V_{t-z}(\widehat{K}_j)\right).
\end{aligned}$$

It follows that

$$\begin{aligned}
&\exp[V_{t+1}(K) + V_{t-1}(K)] = \\
&\sum_{z=0}^{k_j} \sum_{w=0}^{k_j} \binom{t+1}{z} \binom{t-1}{w} \exp((w+z)(U_j + \lambda)) \exp\left[V_{t+1-z}(\widehat{K}_j) + V_{t-1-w}(\widehat{K}_j)\right] \quad (\text{A.18})
\end{aligned}$$

and

$$\exp[V_t(K) + V_t(K)] = \sum_{z=0}^{k_j} \sum_{w=0}^{k_j} \binom{t}{z} \binom{t}{w} \exp((w+z)(U_j + \lambda)) \exp\left[V_{t-z}(\widehat{K}_j) + V_{t-w}(\widehat{K}_j)\right]. \quad (\text{A.19})$$

Consider the sets  $D_L = \{(z, w) : z + w = L\}$ , where  $L$  is an integer,  $0 \leq L \leq 2k_j$ . If one thinks of the  $(k_j + 1) \times (k_j + 1)$  terms in the summations of (A.18) and (A.19) as elements of a  $(k_j + 1) \times (k_j + 1)$  matrix with rows  $z = 0, 1, 2, \dots, k_j$  and columns  $w = 0, 1, 2, \dots, k_j$ , then the set  $D_L$  corresponds to those elements running diagonally from the lower left to the

upper right. It is enough to show that for each of these  $(2k_j + 1)$  sets the terms in equation (A.19) exceed those in equation (A.18). After dividing out the term  $\exp(L(U_j + \lambda))$  in both (A.18) and (A.19), we must show

$$\begin{aligned} & \sum_{l=0}^L \binom{t}{L-l} \binom{t}{l} \exp(V_{t-L+l}(\widehat{K}_j)) \exp(V_{t-l}(\widehat{K}_j)) \\ & \geq \sum_{l=0}^L \binom{t+1}{L-l} \binom{t-1}{l} \exp(V_{t+1-L+l}(\widehat{K}_j)) \exp(V_{t-1-l}(\widehat{K}_j)) \end{aligned} \quad (\text{A.20})$$

For  $L = 0, 1$ , the inequality (A.20) is easily shown by direct calculation coupled with the inequalities of the induction step (A.17). For  $L \geq 2$ , we first observe that the exponential product terms in (A.20) appear in both sums with one exception: the last term ( $l = L$ ) in the right hand sum, i.e.  $\exp(V_{t+1}(\widehat{K}_j)) \exp(V_{t-L-1}(\widehat{K}_j))$ . We can replace this term by its upper bound  $\exp(V_t(\widehat{K}_j)) \exp(V_{t-L}(\widehat{K}_j))$  (see (A.17)) and prove the resulting (stronger) inequality still holds. With this replacement in mind, we combine the coefficients corresponding to identical exponential terms on each side of (A.20). We will use  $a$ 's to represent the combined coefficients for the left hand side of (A.20), and we will use  $b$ 's to represent coefficients for the right hand side.

For  $L$  even ( $L \geq 2$ ), set  $m = \frac{L}{2}$  and define the coefficients

$$a_i = \begin{cases} 2 \binom{t}{L-i} \binom{t}{i} & i = 0, 1, \dots, m-1 \\ \binom{t}{m} \binom{t}{m} & i = m \end{cases}$$

$$b_i = \begin{cases} \binom{t+1}{1} \binom{t-1}{L-1} + \binom{t-1}{L} & i = 0 \\ \binom{t+1}{L+1-i} \binom{t-1}{i-1} + \binom{t+1}{i+1} \binom{t-1}{L-1-i} & i = 1, \dots, m-1 \\ \binom{t+1}{m+1} \binom{t-1}{m-1} & i = m \end{cases}$$

For  $L$  odd ( $L \geq 3$ ), set  $m = \lfloor \frac{L}{2} \rfloor$  where  $\lfloor x \rfloor$  is the largest integer less than  $x$ . Define the coefficients

$$a_i = 2 \binom{t}{L-i} \binom{t}{i} = \quad i = 0, 1, \dots, m$$

$$b_i = \begin{cases} \binom{t+1}{1} \binom{t-1}{L-1} + \binom{t-1}{L} & i = 0 \\ \binom{t+1}{L+1-i} \binom{t-1}{i-1} + \binom{t+1}{i+1} \binom{t-1}{L-1-i} & i = 1, \dots, m \end{cases}$$

For both  $L$  even and  $L$  odd, the coefficients  $a_i$  and  $b_i$ , in their respective (left hand side and right hand side) summations, are the combined coefficients of the exponential term  $\exp(V_{t-i}(\widehat{K}_j)) \exp(V_{t-L+i}(\widehat{K}_j))$  for  $i = 0, 1, 2, \dots, m$ .

Recall that the value  $\exp(V_{t+1}(\widehat{K}_j)) \exp(V_{t-L-1}(\widehat{K}_j))$  in the right hand sum of (A.20) has been replaced with the (larger) term  $\exp(V_i(\widehat{K}_j)) \exp(V_{t-L}(\widehat{K}_j))$  for the calculation of

$b_0$ . This means

$$\sum_{i=0}^m a_i \exp(V_{t-i}(\widehat{K}_j)) \exp(V_{t-L+i}(\widehat{K}_j)) = \sum_{l=0}^L \binom{t}{L-l} \binom{t}{l} \exp(V_{t-L+l}(\widehat{K}_j)) \exp(V_{t-l}(\widehat{K}_j)) \quad (\text{A.21})$$

but

$$\sum_{i=0}^m b_i \exp(V_{t-i}(\widehat{K}_j)) \exp(V_{t-L+i}(\widehat{K}_j)) \geq \sum_{l=0}^L \binom{t+1}{L-l} \binom{t-1}{l} \exp(V_{t+1-L+l}(\widehat{K}_j)) \exp(V_{t-1-l}(\widehat{K}_j)) \quad (\text{A.22})$$

We now show

$$\sum_{i=0}^m a_i \exp(V_{t-i}(\widehat{K}_j)) \exp(V_{t-L+i}(\widehat{K}_j)) \geq \sum_{i=0}^m b_i \exp(V_{t-i}(\widehat{K}_j)) \exp(V_{t-L+i}(\widehat{K}_j)) \quad (\text{A.23})$$

It is straightforward to show that  $a_0 \leq b_0$ . It is also straightforward to show  $a_m > b_m$ . For  $i = 1, \dots, m-1$ , we have

$$a_i - b_i = \binom{t}{L-i} \binom{t}{i} \left[ 2 - \frac{t+1}{t} \left( \frac{i}{L+1-i} + \frac{L-i}{i+1} \right) \right]. \quad (\text{A.24})$$

The function  $g(x) = \frac{x}{L+1-x} + \frac{L-x}{x+1}$  appearing in the bracketed term of (A.24) is strictly decreasing for  $0 \leq x \leq m-1$ . For any given  $t$ , the function  $2 - \frac{t+1}{t}g(x)$  is therefore strictly increasing in  $x$ . Consequently, there is an index  $0 < m^* \leq m$  such that  $a_i - b_i > 0$  for  $i \geq m^*$  and  $a_i - b_i \leq 0$  for  $i < m^*$ . Define

$$\theta_i = \frac{\exp(V_{t-i}(\widehat{K}_j)) \exp(V_{t-L+i}(\widehat{K}_j))}{\exp(V_{t-m^*}(\widehat{K}_j)) \exp(V_{t-L+m^*}(\widehat{K}_j))} \quad i = 0, 1, \dots, m^* - 1$$

and

$$\beta_i = \frac{\exp(V_{t-i}(\widehat{K}_j)) \exp(V_{t-L+i}(\widehat{K}_j))}{\exp(V_{t-m^*}(\widehat{K}_j)) \exp(V_{t-L+m^*}(\widehat{K}_j))} \quad i = m^*, m^* + 1, \dots, m.$$

By the induction hypothesis (A.17), these ratios must satisfy the conditions for  $\theta_i$  and  $\beta_i$

stated in lemma 1. Moreover, by the construction of the coefficients  $a_i$  and  $b_i$  we must have

$$\begin{aligned}\sum_{i=0}^m a_i &= \sum_{l=0}^L \binom{t}{L-l} \binom{t}{l} \\ \sum_{i=0}^m b_i &= \sum_{l=0}^L \binom{t+1}{L-l} \binom{t-1}{l}.\end{aligned}\tag{A.25}$$

Recall the well known Vandermonde Identity (see H.W. Gould, 1956, 1972):

$$\sum_{q=0}^r \binom{x}{q} \binom{y}{r-q} = \binom{x+y}{r}.$$

Applying this to each of the sums in (A.25) we observe

$$\begin{aligned}\sum_{i=0}^m a_i &= \sum_{l=0}^L \binom{t}{L-l} \binom{t}{l} = \binom{2t}{L} \\ \sum_{i=0}^m b_i &= \sum_{l=0}^L \binom{t+1}{L-l} \binom{t-1}{l} = \binom{2t}{L}.\end{aligned}$$

The conditions of lemma 1 are met and therefore  $\sum_{i=0}^{m^*-1} \theta_i a_i + \sum_{i=m^*}^m \beta_i a_i \geq \sum_{i=0}^{m^*-1} \theta_i b_i + \sum_{i=m^*}^m \beta_i b_i$ . Multiplying both sides of this inequality by the term  $\exp(V_{t-m^*}(\widehat{K}_j)) \exp(V_{t-L+m^*}(\widehat{K}_j))$  yields the inequality (A.23) and the theorem is proved. □

## References

- [1] Gould, H.W. 1972. Combinatorial Identities. Morgantown Printing and Binding Co., page 22.
- [2] Gould, H.W. 1956. Some generalizations of Vandermonde's convolution. American Math Monthly, 63, 84-91.

## B Numerical Experiments

We assume there are  $n$  alternatives to choose from, thus permitting maximum variety (i.e., a pack with one unit of each alternative). Without loss of generality, alternative 1 is the consumer's favorite (in expectation), alternative 2 is their second favorite and so on, which means  $U_1 \geq U_2 \geq \dots \geq U_n$ . The indices for the consumer ( $k$ ) and the consumption time ( $t$ ) are suppressed. The cdf for alternative  $i$ 's error  $\epsilon_i$  is  $F_i$  and its density is  $f_i$ .

The probability that the consumer chooses item  $i$  is

$$P(i) = \text{Prob} \{U_i + \epsilon_i \geq U_j + \epsilon_j \quad \forall j, j \neq i\},$$

which implies the marginal choice probability

$$P(i) = \int_{-\infty}^{\infty} \prod_{j \neq i} F_j(U_i - U_j + \epsilon) f_i(\epsilon) d\epsilon. \quad (\text{B.1})$$

We will have frequent occasion to calculate the expected value of the maximum of random variables having the general form  $X_i = c_i + \epsilon_i$  for  $i \in A \subseteq \{1, 2, 3 \dots, n\}$ . The cdf  $G_A(t)$  for  $\max_{i \in A} \{X_i\}$  is

$$\begin{aligned} G_A(t) &= \text{Prob} \left( \max_{i \in A} \{X_i\} \leq t \right) \\ &= \prod_{i \in A} F_i(t - c_i), \end{aligned} \quad (\text{B.2})$$

and so the expected value can be calculated as

$$E \left( \max_{i \in A} \{X_i\} \right) = \int_{-\infty}^{\infty} t G'_A(t) dt. \quad (\text{B.3})$$

The dynamic programming formulation requires that we build the value function for larger packs from the values for smaller packs, e.g., to calculate the value for (2,1,0) we must first know the values for (2,0,0) and (1,1,0). This is accomplished using  $n$  nested loops organized in a particular fashion. The outermost loop corresponds to the favorite alternative (alternative 1), the second outermost loop to the second favorite (alternative 2), and so on. Letting  $k_i$  represent the index for each alternative, the outermost loop uses  $k_1 = 0, 1, 2 \dots, n$ ; the next loop uses  $k_2 = 0, 1, 2 \dots n - k_1$ ; the third loop uses  $k_3 = 0, 1, 2 \dots n - k_1 - k_2$  and

so on. The innermost loop uses  $k_n = 0, 1, 2 \dots n - k_1 - k_2 - \dots - k_{n-1}$ . Calculating the value function for each  $j$ -pack ( $j \leq n$ ) using this nested structure ensures the values for the required smaller packs have been calculated prior to the computation of the larger packs that build upon them. One can show that there are exactly  $\binom{2n-1}{n}$  possible  $n$ -packs, but a total of  $\binom{2n}{n}$  optimal values must be calculated to account for all the smaller sub-packs (0-packs, 1-packs, 2-packs, etc.) used in the valuation build up.

The optimal policy for any  $n$ -pack is simply to select the alternative that maximizes current utility plus expected utility to go. For an  $n$ -pack having alternatives  $i \in A$ , the expected value of following this optimal policy is obtained by calculating the expectation of  $\max_{i \in A} \{V(k_1, \dots, k_{i-1}, k_i - 1, k_{i+1}, \dots, k_n) + U_i + \epsilon_i\}$ , which can be done numerically using equations (B.2) and (B.3) with  $c_i = V(k_1, \dots, k_{i-1}, k_i - 1, k_{i+1}, \dots, k_n) + U_i$ . This is the procedure we coded in MATLAB.

We consider three distributions for the error term: a standard uniform  $U[0, 1]$ , a standard normal  $N(0, 1)$ , and a standard Gumbel, whose cdf is  $F(t) = e^{-e^{-t}}$ . Using standardized distributions imposes no restriction on our experimental outcomes; it merely selects a particular scale (all other scales could be mapped to it), and a particular location, which could be obtained through a suitable shift in each  $U_i$ . These scale and location parameters have no impact on the ordering of  $n$ -pack values.

### 3-Packs

Our previously noted laboratory experiment involving 168 business students (107 BBA, 61 MBA) provided motivation for our first numerical experiment. Students' self-reported selection frequencies (choice probabilities) for their three favorite snacks (favorite, second favorite, third favorite) were obtained via a Qualtrics survey. Half of the students (84) selected choice frequencies that fell into one of six scenarios: (50%, 30%, 20%) (32 students); (40%, 40%, 20%) (15 students); (50%, 25%, 25%) (10 students); (60%, 30%, 10%) (9 students); (40%, 30%, 30%) (9 students); (60%, 20%, 20%) (9 students). (Note: 8 students choose our hypothetical case (33.33%, 33.33%, 33.33%) used in §3.4.) For each of these six scenarios, we calculated parameters  $U_1, U_2, U_3$  that produced the scenario's choice probabilities. For Gumbel distributed errors, this could be done by simply setting  $U_i = \ln(\omega_i)$  where  $\omega_i$  is the scenario's choice frequency for alternative  $i$ . For the uniform and normal distribution, we calculated  $U_1, U_2, U_3$  using equation (B.1) and a simple weighted least squares optimization model where the  $\omega_i$  served as targets. All values are unique up to an additive shift. The

Table 1: Computed Utility Parameters for 3-Packs using Uniform (U) and Normal (N)

Scenario	$U_1(U)$	$U_2(U)$	$U_3(U)$	$U_1(N)$	$U_2(N)$	$U_3(N)$
A (50%, 30%, 20%)	1.4125	1.2857	1.1964	.5202	.1013	-.1982
B (40%, 40%, 20%)	1.0933	1.0933	.9368	.2516	.2516	-.2768
C (50%, 25%, 25%)	1.1457	.9775	.9775	.3835	-.1730	-.1730
D (60%, 30%, 10%)	1.2310	1.0500	.8197	.4900	-.0957	-.8602
E (40%, 30%, 30%)	1.1803	1.1136	1.1136	-.0257	-.2545	-.2545
F (60%, 20%, 20%)	.27345	0.0000	0.0000	.88518	0.0000	0.0000

results are given in Table 1.

There are ten possible 3-packs, and optimal values were computed for each of these ten 3-packs using the three error distributions and the six scenarios. Each combination of a distribution and a scenario is summarized by a vector having ten values, one for each possible 3-pack. While the scales and spacings were different for each vector, there was extraordinary agreement in how the three vectors valued these 3-packs. Indeed, the correlation coefficient between the Gumbel vector and the normal and uniform vectors was over .9985 in all six scenarios, averaging .9995. There were some minor differences. In scenarios B, C and D and F, the Gumbel included several ties that were not obtained using the normal and uniform distribution. In scenario B, the Gumbel produced a three-way tie between (2,1,0), (1,2,0) and (1,1,1) for the highest value, whereas the normal and uniform produced a two-way tie between (2,1,0) and (1,2,0) and assigned (1,1,1) to the third highest value. Similarly, in scenario C, the Gumbel produced a three-way tie between (2,1,0), (2,0,1) and (1,1,1) for the highest value, whereas the uniform and normal produced a two-way tie between (2,1,0) and (2,0,1) for the highest value and assigned (1,1,1) to the third spot. In scenario D, the Gumbel produced a tie between (2,0,1) and (1,1,1) for the 4th spot, whereas the uniform and normal split these into (2,0,1) (4th spot) and (1,1,1) (5th spot). In scenario F, both the Gumbel and normal produced a two-way tie between (2,0,1), (2,1,0) for the top spot and assigned (3,0,0) to the third spot, whereas the uniform choose (3,0,0) for the top spot with (2,1,0) and (2,0,1) tied for the second spot. All of the aforementioned discrepancies involved minute differences in the value functions (approximately 1.2% of the range in values), but it was more than could be attributed to numerical error. Overall, values obtained using the Gumbel distribution were highly representative of what we would expect to obtain for the uniform and normal error distributions.

## 6-packs

We repeated the foregoing experiment on 6-packs. Since we did not have experimental data (choice frequencies) from students in this case, we simulated six choice frequencies using six random draws  $X_i$  from a uniform distribution on  $[0,1]$ . The choice frequencies  $f(i)$  were then calculated using  $f(i) = \frac{\prod_{l=1}^i X_l}{\sum_{l=1}^6 \prod_{l=1}^i X_l}$ , which ensured  $f(1) \geq f(2) \geq \dots \geq f(6)$  and  $\sum_{i=1}^6 f(i) = 1$ . One hundred scenarios were simulated, and these ran the gamut from broadly distributed (24%, 18.3%, 17.2%, 16.8%, 14.4%, 8.9%) all the way to highly skewed (98% for choice 1, 1.9% for choice 2). In contrast to the previous experiment, many of these scenarios had a high number of low frequency alternatives (63 scenarios had at least one choice frequency below 1%).

As before, the  $f(i)$  for each scenario became targets in an optimization framework to calibrate utility parameters (for the uniform and normal) so that each distribution's choice probabilities matched the  $f(i)$  to within 5 digits. These utility parameters were then used to calculate the values for all 462 possible 6-packs under each error distribution, resulting in a 462 dimensional vector of values for each scenario and each distribution. The average correlation between the Gumbel valuation vector and the uniform valuation vector slipped to .971, whereas the correlation between the Gumbel and normal valuation vectors remained nearly the same at .992. Upon examining the results, it became apparent that the results were impacted by the large number of scenarios having one or more low frequency alternatives. These scenarios produced notably lower correlations, the lowest of which, scenario 94, produced a .851 correlation between the Gumbel and uniform valuations (but .943 between the Gumbel and normal). In scenario 94, the three lowest choice frequencies were .0003, .000049, and .0000007. In the Gumbel model, where the utilities are calculated using the formula  $U_i = \ln(f(i))$ , low choice frequencies are mapped into much lower utilities and thus very low valuations for packs that include them (note that if any  $f(i) \rightarrow 0$ , the Gumbel valuation goes to  $-\infty$ ). In contrast, the utility parameters for the uniform distribution are bounded in an interval of length 1 and all of its 6-pack valuations are bounded as well. This means packs including low frequency alternatives are compressed by the uniform model and receive higher valuations compared to the Gumbel model (and, as might be expected, the normal model). If valuations are limited to packs consisting of higher probability alternatives, these value distortions disappear. For example, in scenario 94, if one eliminates those packs including one or more of the three lowest choice frequencies, the correlation between the remaining Gumbel and uniform valuations jumps to .995. In sum, we find that valuations for packs that are likely to be selected (i.e., attractive to a consumer) remain very highly correlated in all three models. Distortions between the Gumbel and uniform (and between the normal and uniform) occur in packs including one or more very low probability

alternatives. Since optimization of  $n$ -packs implicitly ignores such packs, we would expect these distortions, when they occur, to have no meaningful impact on our results.