1 Introduction

A great deal of research has focused on two key consumer decisions: (1) the shopping, or purchase, decision which is usually made in store; (2) the consumption decision which is made later at the time of consumption. For consumer packaged goods, the two decisions are inextricably linked even though they occur at different times and in different places.

Shopping decisions are made for future consumption, which generally occurs over multiple consumption occasions. On each such occasion, only products selected previously while shopping are available to be consumed. The shopping decision therefore creates a set of products for future consumption and so is inherently forward-looking. At the same time, the shopping decision the product alternatives that are available for future consumption compared to the full assortment in store. Each successive consumption decision may further constrain the product alternatives available, depending upon how many units of each product alternative were selected when shopping. If the consumer has only a single unit of a particular product alternative remaining, then consuming it would preclude choosing that alternative on all succeeding consumption occasions. Because consumption decisions, like the shopping decision, affect the expected utility of future consumption, these decisions are

people prefer to preserve options for the future, even when doing so leads to less desirable outcomes. Our canonical model strongly supports this finding. The experiments reported in their study involve two-stage choices, where only a single item is chosen in the second stage. Similar two-stage choice models have been applied to consideration set formation (Hauser and Wernerfelt 1990, Roberts and Lattin 1991) and to choice among retail assortments (ahn and Lehmann 1991). Like the models we propose herein, these two-stage models specify Gumbel-distributed errors to represent preference uncertainty.

Guo (2010) developed a structural econometric model for consumers' choice of assortments (-packs). His model allows for consumption flexibility, due to future preference uncertainty as well as state dependence; our model addresses only the former. Guo estimated his model on scanner panel data for yogurt purchases. Because consumption data was not available, Guo estimated the consumer's valuation of each assortment (what we call -packs) using simulation. This involved simulating error streams for each alternative over the consumption horizon and assuming the consumer selects the alternative offering the highest utility on each consumption occasion. We note that such a consumption policy is plausible but not optimal. Guo found that allowing for both future preference uncertainty and state dependence offers better in- and out-of-sample fits for the scanner panel data than more restricted nested models. However, his parameter estimates indicate positive state dependence—this is the opposite of variety-seeking, which is received wisdom in consumer psychology (Simonson 1990, Read and Loewenstein 1995). Guo also found that consumers make consistent multi-product purchases; that is, they purchase horizontally-varied sets of products but purchase similar sets of products over time. In an earlier study, Guo (2006) determined that consumption flexibility, due to preference uncertainty, also affects firm decisions about product variety and pricing. sing a duopoly model, Guo identified the conditions under which consumers purchase multiple competing products. He found that, if consumers have relatively homogeneous preferences, firms can actually make lower profits by falling into a "flexibility trap" by pricing to attract primary demand.

The work that is closest to ours is due to Walsh (1995). In this paper, the author modeled consumption decisions for assortments with two product alternatives. Both alternatives' future utilities are random, and the problem reduces to an equivalent one in which one alternative has random utility and the other has constant utility (a reduction that only works for assortments with precisely two alternatives). Assuming that consumers are forward-looking, Walsh developed dynamic equations that describe optimal consumption behavior and the associated value function. Although the form of the policy and the value function are not available in closed-form, Walsh's analysis yielded three interesting findings: (i) consumers may not choose the alternative offering the greatest utility on a particular consumption occasion; (ii) more inventory of an alternative makes it more likely to be selected; (iii) adding an additional unit to the assortment causes the utility of that assortment to increase by more than the expected utility of the item added. Our canonical model generalizes Walsh's findings (and adds some refinements) while enabling normative predictions for shopping decisions. Further, our generalized model (including an outside option for consumption) demonstrates that the canonical model represent a boundary solution. Compared to Walsh's model, ours (i) apply to -packs of any size and with any number of product alternatives; (ii) result in a closed-form value function that can be maximized to determine each consumer's optimal -pack; (iii) are based on marginal choice probabilities and so can be customized to individual consumers and used for decision support. The tradeoff we make is in using the multinomial logit framework (deterministic utility plus Gumbel-distributed errors) to describe future utilities; Walsh used a general error distribution. Given the ubiquity of the multinomial logit in discrete choice and assortment planning models, we feel that this tradeoff is justified.

Another related vein of research involves assortment optimization in the revenue management literature. The multinomial logit (MNL) plays a prominent role in this research. One of the earliest papers in this vein is due to van Ryzin and Mahajan (1999), who used MNL embedded in the demand model of a newsvendor problem and derived optimal profit functions under several reasonable assumptions. The authors showed that the profit-maximizing assortment is some subset of the most popular products (the most popular products have the highest probability of being selected). The authors used the concept of majorization to derive sufficient conditions that ensure the profits of one category dominate those of another. In a subsequent paper, Talluri and van Ryzin (2004) introduced a dynamic model and developed conditions on the choice probabilities that ensure the optimal assortment is some contiguous set of the highest fare products (the "nested by fare order" property). They developed necessary and sufficient conditions that once again involve the concept of majorization, and the MNL choice model was shown to satisfy these conditions. More recently, Rusmevichientong and Topaloglu (2012) showed that these results remain valid for MNL in the presence of parameter uncertainty (for the choice probabilities) and a capacity constraint.

In contrast to these papers, our model addresses assortment optimization at the consumer level and not the retailer level. Moreover, our model does not use MNL choice probabilities to capture consumer demand, but instead uses the random utility of MNL to capture preference (utility) fluctuations for a given consumer over time. Our model also focuses on the combination of alternatives and that comprise a consumer's optimal n-pack. Majorization plays an important role in our work as well, but we do not use it as an assumption to prove a theoretical result. Rather, we find that a stronger form of majorization, what we have called "strong majorization," characterizes the relationship between the optimal solutions of our two main models.

3 Expected Utility of an n-Pack: The Canonical Model

3.1 Assumptions

Consistent with the extant literature, we begin by assuming (in this section) that the consumer selects an alternative from a preselected -pack on each consumption occasion. There are M distinct product alternatives available in the product category (the full assortment available in store) although only m alternatives are represented in the -pack (m n; m M). The utility parameters for each alternative are U_i ($i = 1; 2; :::; M$). These parameters could be a function of many things; however, we take them to be fixed for ease of exposition. On any particular consumption occasion t, the utility that consumer j receives from a particular alternative i is U_{ji} + $_{jit}$ where the random errors $_{jit}$ are assumed to be independent Gumbel distributed with CDF $F(z) = \exp(\; e^{(z \quad)=})$. The errors account for a variety of unmodeled factors that affect consumption decisions, and each consumption occasion t represents a fresh draw for these errors. For example, a consumer might prefer vegetable soup on most consumption occasions but prefer chicken soup when they are feeling ill—this would be captured in the error term. Like Walsh (1995) and Guo (2010), we assume these errors become known to the consumer at the time of consumption but not before. Given the canonical model's assumption that one unit is consumed per period, we must have $t = 1; 2; : : : ; n$ periods in the consumption horizon.

Without loss of generality, we may assume that the problem has been normalized so that the errors are Gumbel with $= 0$ and $= 1$ (observe that U_{ji} + $_{jit}$ U_{jl} + $_{jlt}$ if and only if $\frac{1}{}U_{ji} + \frac{1}{}(\;_{jit}\quad)$ $\frac{1}{}U_{jl} + \frac{1}{}(\;_{jlt}\quad)$, but $\frac{1}{}(\;_{jit}\quad)$ is standard Gumbel for all $(j; i; t)$). The expectation of a standard Gumbel is $E(\;_{jit}) = \;_{0}^{1} \ln(z)e^{\;z}dz$; this is Euler's constant and denoted by . The expected utility of each product is therefore $E(U_{ji} + \;_{jit}) =$ U_{ji} + : Without loss of generality, alternatives are oi5it6h5o46-418(Lik)28(e35-307.)-25TJ/7686TJ/7686T by $V(k_1; k_2; \ldots; k_M)$.

A timeline that illustrates the dynamics for consuming a 3-pack consisting of one unit of and we would (strictly) prefer alternative 2 if the inequality were reversed. Ties can be

Moreover, by the definition of $a(k_1; k_2)$ (see (1))

$$
\begin{aligned}
a(k_1; k_2) &= U_1 - U_2 + V(k_1 - 1; k_2) - V(k_1; k_2 - 1) \\
&= \ln \frac{n - 1}{k_1 - 1}
\end{aligned}
$$

future preference uncertainty.[2] The term $\ln(n!)$ captures the effect of an -pack's size while the term $\ln((k_1)!(k_2)!\quad(k_M)!)$ captures the effects of both variety and inventory. For an -pack with a fixed number of units n, the choice premium is increased by including more alternatives and/or "flattening" the distribution of alternatives (k_i). The maximum choice premium is $\ln(n!)$, which is realized when there is exactly one unit of distinct alternatives; the minimum choice premium is 0, which is realized when $k_i = n$ for some alternative . For any -pack then, the ratio of the choice premium to $\ln(n!)$ can be interpreted as the proportion of the available choice premium captured by that -pack.

The intuition behind this optimal policy can be made clear by considering a simplified case. Let us suppose a given consumer has a 2-pack consisting of one unit of alternative 1 (their favorite), and one unit of alternative 2 (their second favorite), with $U_1 > U_2$. On the first consumption occasion, suppose the observed error terms are e_1 and e_2 with $e_2 > e_1$. Even if $U_2 + e_2 < U_1 + e_1$, alternative 2 represents the better consumption choice This is because U_1 and U_2 are fixed, and so the of the realized errors (one now, one later) will ultimately decide the total utility received over both consumption occasions. Because

$$
\begin{aligned}
\mathrm{Prob}(i) &= \mathrm{Prob}\{U_i + V(k_1; \; ; k_{i-1}; k_i \; 1; k_{i+1}; \; ; k_M) + \; _i \\
&\qquad U_j + V(k_1; \; ; k_{j-1}; k_j \; 1; k_{j+1}; \; ; k_M) + \; _j \quad \forall j \neq i\} \\
&= \mathrm{Prob}(V(k_1; \; ; k_M) + \ln(k_i) + \; _i \quad V(k_1; \; ; k_M) + \ln(k_j) + \; _j \quad \forall j \neq i) \\
&= \mathrm{Prob}(\ln(k_i) + \; _i \quad \ln(k_j) + \; _j \quad \forall j \neq i) \\
&= \frac{k_i}{\sum_{l=1}^{M} k_l}
\end{aligned}
$$

The last equality follows from the standard logit probability formula with the customary utility parameter "U_i" replaced by $\ln(k_i)$.

3.3 Identifying a Consumer's Optimal n-Pack

Given a consumer's U_i (as can be estimated from purchase histories or using preference elicitation methods), the value function in (4) can then be optimized over all possible integer quantities $(k_1; k_2; \ldots; k_M)$ ($k_i \quad 0; \sum_{i=1}^{M} k_i = n$) to obtain the consumer's , $(k_1; k_2; \ldots; k_M)$. The optimal pack represents the solution to the first stage (the selection/shopping stage) of our two-stage problem (selection and then consumption). Figures 2a and 2b show the optimal -packs of sizes $n = 2$ and $n = 3$. The optimal -packs vary by region, depending on in their ranked utilities; this is because translating all utilities by a constant translates all -pack values by a constant as well. We assume that the utilities are ordered so that $U_1 \quad U_2 \quad U_3$, fixing $U_1 = 0$ for identification purposes. For $n = 2$, the distribution of product utilities is captured in the difference $U_1 - U_2$

Figure 2a. Optimal -pack by Region for n = 2.

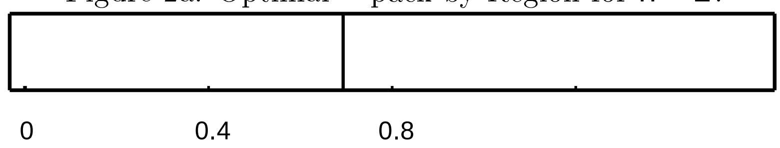
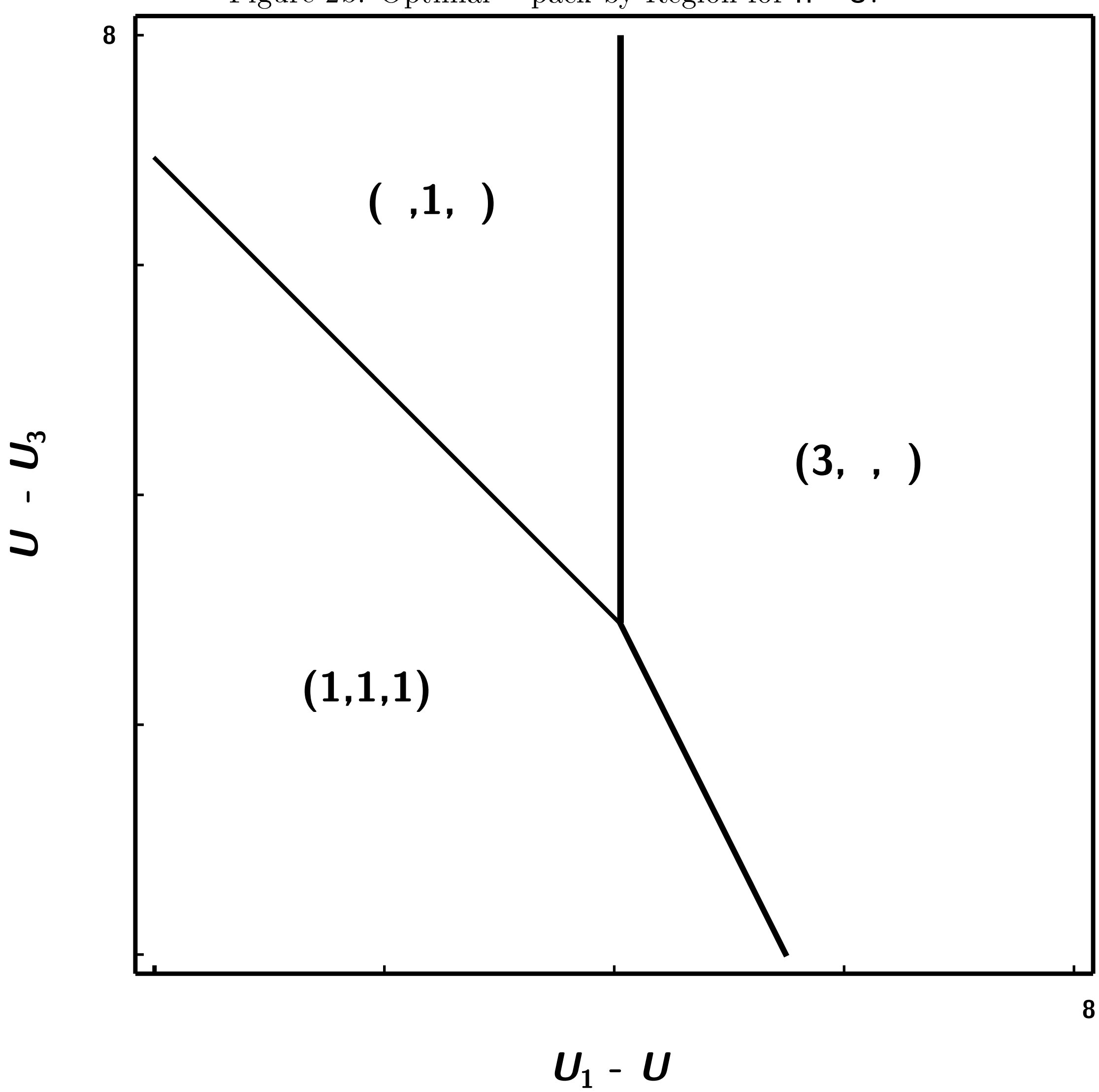


Figure 2b. Optimal -pack by Region for n = 3.



many solvers, and we experienced significant difficulties trying to solve this using off-the-shelf software (in large part because many of the k_i are precisely zero and the optimizer would terminate if these slipped into negative territory). Fortunately, it is quite easy to solve this problem using a simple optimization algorithm based on swaps that is guaranteed to produce a global optimal solution in at most swaps. This algorithm is described next.

Suppose we have an incumbent feasible solution $k_1; k_2; \ldots; k_M$ ($\sum_{i=1}^{M} k_i = n$, k_i 2 N^0) and we want to improve it. One way is to increase a non-negative k_i by one unit and decrease a currently positive k_j by one unit. The net change in the objective function caused by this one unit swap is

$$U_i \quad \ln(k_i + 1) \quad U_j + \ln(k_j); \tag{5}$$

and this suggests the following greedy technique. Interpreting $\ln(0) = \lim_{a! \ 0^+} (a) = \ 1$, calculate the optimal indices

$$i \quad = \quad \underset{i}{\operatorname{argmax}} (U_i \quad \ln(k_i + 1) \; ; \; i = 1; \ldots; M) \tag{6}$$

$$j \quad = \quad \underset{j}{\operatorname{argmin}} (U_j \quad \ln(k_j) \; ; \; j = 1; \ldots; M) : \tag{7}$$

Ties in the maximum can be broken by selecting the alternative with the smallest index; ties in the minimum can be broken by selecting the alternative with the largest index. Then define the difference

$$U_i \quad \ln(k_i \ + 1) \quad U_j \ + \ln(k_j \ ) : \tag{8}$$

If > 0 then increase k_i by one unit and decrease k_j by one unit; if no such combination exists, then stop. Observe that if > 0, the objective function increases by a strictly positive amount after each swap. Otherwise, 0 and we must have achieved the stopping condition

$$\underset{i}{\operatorname{Max}} (U_i \quad \ln(k_i + 1) \; ; \; i = 1; \ldots; M) \quad \underset{j}{\operatorname{Min}} (U_j \quad \ln(k_j) \; ; \; j = 1; \ldots; M) : \tag{9}$$

Theorem 2. (Optimality of the Swapping Algorithm)
$(k_1; k_2; \ldots; k_M)$ with $\sum_{i=1}^{M} k_i = n$ k_i 2 N^0

optimal, and they can be exploited in several ways. One of these is the connection between the optimal (n)-pack and the optimal $(n+1)$-pack, which is described next.

Theorem 3. $(k_1; k_2; :::; k_M)$ n

$(n + 1)$ i

U_i $\ln(k_i + 1)$

The latter theorem not only means we can build larger optimal packs from smaller optimal packs, but it also identifies .

For a retailer looking to encourage a consumer to buy an additional unit, this information would be extremely useful.

The optimality conditions can be used to obtain insights into the structural properties of the optimal simply "sequential choice") as a benchmark for measuring variety. Sequential choice implies the consumer is allowed to choose any alternative from the full assortment (i.e., every alternative in a category that could be selected from a store) on each consumption occasion. This wait-and-see approach means the consumer can observe the random component of utility ($_i$) for every alternative in the full assortment immediately before making a consumption decision. The consumer thus maximizes their utility on every consumption occasion and cannot obtain any greater utility than this when consuming items on consecutive occasions. We show next that the optimal -pack selected is the same as the most probable set of items chosen sequentially.

As before, we describe each -pack using an M

We then observe that

$$(x_1; x_2; \ldots; x_M$$

3-pack with exactly one alternative; 80 students had an optimal 3-pack with exactly two alternatives; and 62 students had an optimal 3-pack with exactly three alternatives. sing the same self-reported choice frequencies, we found that in 102 of the 168 cases (60.7%), the probability of a student consuming less variety than their optimal 3-pack in a sequential choice experiment would be greater than their probability of consuming more. Additionally, the average probability of a student consuming less variety than their optimal 3-pack over all 168 cases was computed to be .383, whereas the average probability of consuming more variety was computed to be .151. This offers additional support for the conjecture that consuming less variety in sequential choice experiments may be a consequence, in part, of probabilistic principles stemming from rational decision-making. Additional work is under way to rigorously test this and other conjectures regarding variety.

3.5 Robustness of the Model: Assessing the Impact of the Gumbel Assumption

To assess the robustness of our results to other error distributions, we conducted several numerical experiments to ensure that our results were not overly dependent on the assumption of a Gumbel distribution. We provide a summary of the results here; the reader is referred to Appendix B for the details.

Two additional error distributions were selected, the uniform and the normal. In the first numerical experiment, we analyzed 3-packs based on utilities calculated from actual choice data. In the second experiment, we analyzed 6-packs based on utilities calculated from simulated choice probabilities. In total, 106 test problems were analyzed, six involving 3-packs and 100 involving 6-packs. In general, there were virtually no meanimgful discrepancies in valuations of -packs . Consequently, the error distribution appears to have little if any impact on a consumer's valuation of their most preferred -packs (say the ratings for their top 10-20 -packs). There were some discrepancies in valuations for problems that included alternatives the consumer would rarely (if ever) select, i.e., test problems that included one or more "unpopular alternatives" with choice probabilities approaching 0. In such cases, the corresponding utility in the normal and Gumbel models becomes unbounded from below, whereas the utility in the uniform model is always bounded. For this reason, the values for the Gumbel and normal distributions tended to track each other closely for all -packs, whereas the values for the uniform tended to diverge for those -packs that included unpopular alternatives, which are inherently "low-value" -packs. Because our analysis is based on determining a consumer's optimal -pack, differences in low-value -packs have no impact on our results.

4 Expected Utility of an n-Pack: The Generalized Model

4.1 The Optimal Value Function

In this section we assume the consumer may select the outside option on any consumption occasion and thus reject all items remaining in their pack. The introduction of an outside option effectively allows for different consumption rates. It does, however, add complexity compared to the canonical case.

We will again use n to denote the number of total units in the -pack, and M to denote the total number of distinct alternatives available. Let $k_i \in N^0 = \{0, 1, 2, 3, \dots\}$ represent the number of units of alternative i in the -pack. The utility parameter for each alternative is denoted by U_j for $j = 0, 1, \dots, M$, (note that we include U_0, the utility of the outside option). The number of consumption occasions is denoted by t, which is also the number of time periods in our dynamic analysis, and the value function with t periods to go is denoted by $V_t(k_1, k_2, \dots, k_M)$, which means consumption periods in the generalized model are numbered (as is frequently done in dynamic programming models). The value function in the terminal (salvage) period (period 0) is $V_0(k_1, k_2, \dots, k_M) = 0$. One can think of this as an n-pack becoming worthless if its expiration date is reached without having been consumed.

The "no consumption" option is represented by the subscript 0, and we can represent the set that is ultimately consumed (or "realized") after t consumption occasions by an $M + 1$ dimensional $(x_0, x_1, ; x_M)$r0Td[(i)]T02Tf4.5530Td[7x lead to

$$
\begin{array}{rl}
\text{(four periods)} & I_4(4,2,1) = \{(4,0,0),\ (3,1,0),\ (3,0,1),\ (2,1,1),\ (2,2,0),\ (1,2,1)\} \\
\text{(three periods)} & I_3(3,2,1) = \{(3,0,0),\ (2,1,0),\ (2,0,1),\ (1,1,1),\ (1,2,0),\ (0,2,1)\} \\
\text{(two periods)} & I_2(2,2,1) = \{(2,0,0),\ (1,1,0),\ (1,0,1),\ (0,1,1),\ (0,2,0)\} \\
\text{(one period)} & I_1(1,2,1) = \{(1,0,0), (0,1,0), (0,0,1)\}
\end{array}
$$

Observe that in each of these three sets, the first input (y_0) is taken to equal the number of consumption occasions. This is appropriate since y_0 is the upper bound on the number of times the "no consumption" option could be invoked, which is equal to the number of consumption periods, . While the number of terms can be quite large, it is bounded independently of the number of consumption occasions t. Indeed, there are at most $\prod_{i=1}^{M}(k_i+1)$ elements in $I_t(t, k_1, k_2, \ldots, k_M)$, which corresponds to the number of distinct subsets of $(k_1, k_2, \ldots, k_M)$ padded by the appropriate number of "outside option" selections to bring the total number of selections to t. This upper bound is obtained for all t n.

Theorem 5. (Optimal Value of an n-Pack, Generalized Model)

or

t

$$V_t(k_1, k_2, \ldots, k_M) = \ln\left[\sum_{(x_0, x_1, \ldots, x_M) \in I_t(t, k_1, k_2, \ldots, k_M)} \frac{t!}{x_0! x_1! \cdots x_M!} e^{\sum_{j=0}^{M} x_j U_j}\right] + t \qquad (13)$$

The optimal policy at each consumption occasion is to select, among the available alternatives, the one that maximizes current utility plus expected utility-to-go, i.e,. the one that maximizes $U_0 + {}_0 + V_{t\ 1}(k_1, k_2, \ldots, k_M)$; $U_j + {}_j + V_{t\ 1}(k_1, k_2, k_j\ \ 1, \ldots, k_M)$ for $k_j > 0$. nlike the canonical version, there is no additional simplification in the optimal policy.

While the value function is somewhat complicated, it can be simplified under the assumption t n, which we would expect to hold in practice. As noted earlier, there are a constant $\prod_{i=1}^{M}(k_i + 1)$ terms in the summation of (13). Additionally, defining the shifted parameters $U_i^0 = U_i\ \ U_0$ (so that the outside option has utility parameter $U_0^0 = 0$), the value function (13) can be expressed as

$$V_t(k_1; k_2; :::; k_M) = \ln \left[\sum_{x_i \; k_i ; i \; 1} \frac{t!}{(t \; \sum_{i=1}^{M} x_i)!} \frac{e^{\sum_{j=1}^{M} x_j U_j^0}}{x_1! \; x_M!} \right] + t(\; + U_0): \qquad (14)$$

This simplified form is easier to manipulate and is used extensively in Theorem 7.

The terms in the value function (13) generalize the probability interpretation established for the canonical model in section 3.4. There, we established that the value function for the -pack $(k_1; k_2; :::; k_M)$ in the canonical model could be equated to the log-probability of consuming the same set of products in a sequential choice experiment; i.e., an experiment in which the consumer can select the product from the full assortment that is most preferred on ea47.9701f6(consupation)-f6(is)-2prob7927Tduea47.9n,eib792 m7alub792ut97(of)]TJ-maximiz

4.2 Variety, Consumption Horizons, and the Outside Option

Imagine two consumers, A and B, both of whom like exactly three types of wine: Chardonnay, Merlot, and Cabernet. Both consumers prefer Chardonnay 70% of the time, Merlot 15% of the time, and Cabernet 15% of the time. However, Consumer A enjoys a bottle every evening whereas Consumer B enjoys a bottle about once a week. If we assume the opportunity to consume wine presents itself every evening, then Consumer A has a very small value for U_0 and thus a high usage rate for wine whereas Consumer B has a much higher value for U_0 and thus a low usage rate for wine. Given this information, which 3-pack of wine should each consumer buy? (We ignore, of course, the fact Consumer A would probably tend to buy a larger pack size.)

The fact that consumer A always chooses wine is evidence of an intrinsically low value for U_0, one that is exceeded by U_1, U_2, and U_3

Table 2. Optimal 3-pack for the Wine Example.

	Horizon (t)							
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

captures what is typically known in dynamic programming as the "end of horizon effect." Over slightly longer horizons, carrying greater variety than $(k_1; k_2; :::; k_M)$ would still be optimal to address these end of horizon effects. Nevertheless, end of horizon effects typically dissipate over sufficiently long horizons, and this dynamic model is no different. As the horizon lengthens, there are sufficient opportunities to match the right item with the right consumption occasion, and the optimal -pack converges to that of the canonical model.

The relationship between the optimal -pack in the generalized model and the optimal -pack in the canonical model can be characterized more precisely. The precise result is stated in the following theorem (where alternatives are again labelled so that $U_1 \quad U_2 \quad\quad U_M$). The condition $t \quad n$ is included to simplify the proof; otherwise, the number of terms in the value function would also depend on t.

Theorem 6. $t \quad n$
$(k_1; k_2; :::; k_M)$ $(q_1; q_2; :::; q_M)$
$q_j > k_j$ $q_i < k_i$ $j < i$

The theorem requires that the components of the optimal vectors $k = (k_1; k_2; :::; k_M)$ and $q = (q_1; q_2; :::; q_M)$ follow a strict pattern. Assuming the vectors are not identical, then there is an index, say c, where a change occurs; for $i \quad c$ we must have $k_i \quad q_i$, and for $i > c$

is weaker than the condition posed in Theorem 6. Consider our previous example involving $k = (5; 2; 2; 0; 0)$ and $q = (3; 3; 1; 1; 1)$; it is clear that $k \quad q$ even though this pair does not satisfy the relationship described by the theorem. The contrast between the two concepts can be sharpened by looking at the difference in partial sums; whereas majorization requires that $\sum_{i=1}^{l} x_{(i)} \quad \sum_{i=1}^{l} y_{(i)} \quad 0$ for $l = 1; 2; \ldots; M$, the condition $k_i \quad q_i$ for $i \quad c$ and $k_i \quad q_i$ for $i > c$ requires that $\sum_{i=1}^{l} x_{(i)} \quad \sum_{i=1}^{l} y_{(i)}$

The property of diminishing marginal value makes intuitive sense. Because the optimal value is based on matching the product with the consumption occasion, adding additional periods should not yield proportional gains in value. Adding additional periods means we are searching for better opportunities in the right hand tail of the error distribution, and better payoffs should become increasingly difficult to obtain as time increases. Dynamic models in revenue management often require this type of structure, and so the property is an important one if the model is to find additional applications in this area.

5 Summary and Future Research

We have proposed a utility-maximizing model based on consumers' long-run consumption preferences to estimate the value they can expect to receive from an -pack of substitutable products. Our canonical model predicts that (i) strategic consumers will choose different product alternatives in proportion to their available inventory and (ii) the total value consumers derive from an -pack increases in the pack's utility parameters but decreases as the distribution of products within the -pack becomes more concentrated. This result could explain the seemingly excessive variety that has been observed in behavioral experiments on -pack selection for future consumption. Our generalized model demonstrates that the inclusion of an outside option (effectively reducing the consumption rate), which to our knowledge has never been done, would lead to even more variety in -pack alternatives and even greater dispersion in -pack quantities.

Our model assumes that both the consumption utilities U_i and the distribution of stochastic errors $_i$ are stationary, but this might not always be the case. Allowing for non-stationarity in consumption utilities (i.e., variety-seeking, state dependence) or in the stochastic error distribution (i.e., learning) might lead to new results and insights. Our model also assumes that there is no discounting of future utilities, so temporal discounting is another possible area for future research. The direct approach, introducing a discount factor on the expected "value-to-go" function, sacrifices the virtue of a closed-form value function. However, it may still be possible to analyze the value function implicitly or to introduce a different discount mechanism that preserves the closed-form solution. Our model also does not currently allow for any replenishment of inventory by the consumer. Could replenishment be included as yet another choice at each consumption occasion? This would allow erations. Perhaps the most obvious extension is to assume that the consumption utility parameters U_i are functions of product attributes, such as price. Still another useful extension would be to investigate how -pack valuation affects consumers' willingness to pay. For example, many retailers implicitly offer the option to purchase an -pack including only a single product alternative at a low price per unit, or purchase single units of different product alternatives at a higher price per unit. For example, a 6-pack of a single brand/type of beer might cost $8.99 ($1.50/beer) while purchasing different beers individually might cost $1.99/beer. Depending on the difference between consumers' valuation of their optimal -pack and an -pack with only their favorite alternative, the retailer may be able to price a "build your own 6-pack" option to extract additional revenues while also increasing consumers' utility. Pricing -packs and designing promotions that target individual consumers are natural applications for this type of model.

As noted in the introduction, there are some applications of our model that require very little additional work. One such application involves auctioning a set of related products.

model to include a disutility term for storage and/or the price of units, and the optimization would then determine the appropriate pack size along with the optimal pack. Given the simple structure of the value function in the canonical model, the optimization problem thus created might be relatively straightforward to analyze.

References

[1] Bown, N. J., D. Read and B. Summers. 2003. The lure of choice. Journal of Behavioral Decision Making, 16(4), 297-308.

[2] Gould, H.W. 1972. Combinatorial Identities. Morgantown Printing and Binding Co., page 22.

[3] Gould, H.W. 1956. Some generalizations of Vandermonde's convolution. American Math Monthly, 63, 84-91.

[4] Guo, L. 2010. Capturing consumption flexibility in assortment choice from scanner panel data. Management Science, 56(10), 1815-1832.

[5] Guo, L. 2006. Consumption flexibility, product configuration, and market competition. Marketing Science, 25(2), 116-130.

[6] Hauser, J. R., and B. Wernerfelt. 1990. An evaluation cost model of consideration sets. Journal of Consumer Research, 16(4), 393-408.

[7] ahn, B. E., and D. R. Lehmann. 1991. Modeling choice among assortments. Journal of Retailing, 67(3), 274.

[8] ahneman, D., and J. Snell. 1990. Predicting utility. Insights in decision making: a tribute to Hillel J. Einhorn. Robert M. Hogarth (ed). niversity of Chicago Press, Chicago, IL, 295-310.

[9] reps, D. M. 1979. A representative theorem for "preference for flexibility". Econometrica (Pre-1986), 47(3), 565.

[10] March, J. G. 1978. Bounded rationality, ambiguity, and the engineering of choice. Bell Journal of Economics, 9(2), 587.

[11] Pessemier, E. A. 1978. Stochastic properties of changing preferences. The American Economic Review, 68(2), 380.

[12] Read, D., and G. Loewenstein. 1995. Diversification bias: Explaining the discrepancy in variety seeking between combined and separated choices. Journal of Experimental Psychology: Applied, 1(1), 34-49.

Appendix A (Proofs)

THEOREM 2

Consider an -pack with quantities k_i for alternative i that satisfies the stopping rule (9). Then for any other pack with quantities k_i^0, we can define two sets: $I^+ = f i : k_i > k$

$j \; 2 \; I \;\; ; m_j \; 2 \; N; 1 \quad m_j \quad k_j^0 \quad k_j$ from (18). These pairings can be done in any manner. We can thus rewrite (15) using a single summation where each (paired) term has the general form

$$
\begin{gathered}
U_i \quad U_j \quad \ln(k_i \quad l_i) + \ln(k_j + m_j) \\
i \; 2 \; I^+; l_i \; 2 \; N^0; l_i \quad k_i \quad k_i^0 \\
j \; 2 \; I \;\; ; m_j \; 2 \; N; 1 \quad m_j \quad k_j^0 \quad k_j
\end{gathered}
\tag{19}
$$

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

$$
\begin{aligned}
& U_i \quad \ln(k_i) \quad U_j + \ln(k_j + 1) \qquad 0 \text{ all } i; j \\
=) \quad & U_i \quad \ln(k_i \quad l_i) \quad U_j + \ln(k_j + m_j) \qquad 0 \text{ all } i; j
\end{aligned}
$$

which implies all the terms in (19) are non-negative and thus so is the summation in (15). Thus $V(k_1; k_2; :::; k_M) \quad V(k_1^0; k_2^0; :::; k_M^0) \quad 0$ as was desired.

To show the algorithm converges in at most 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 $U_i \quad i \; "$ where " 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 index, and breaking ties in (7) by selecting the alternative with the 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 $g \; 6 \; l$. We must have

$$
U_g \quad g \; " \quad \ln(k_g + 1) \; > \; U_i \quad i \; " \quad \ln(k \ln
$$

The gain (4) in our objective function for the current swap is

$$4 = U_g \quad g \;" \quad \ln(k_g + 1) \quad U_l + l \;" + \ln(k_l) \; > \; 0: \tag{22}$$

The alternative g cannot be the minimizer for (7)–and hence the loser–in the 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 $k_l \quad 1$ units for alternative l, we may apply (5) to calculate the possible gains in the objective function at the next iteration as

$$U_i \quad i \;" \quad \ln(k_i + 1) \quad (U_g \quad g \;" \quad \ln(k_g + 1)) \qquad i \not\in l; g \tag{23}$$

$$U_l \quad l \;" \quad \ln(k_l) \quad (U_g \quad g \;" \quad \ln(k_g + 1)) : \tag{24}$$

The potential gains in (23) are all negative by (20); the potential gain in (24) is negative by (22). 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^0 ($g^0 \not= g$), 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 $(k_g + 1)^{st}$ unit). For case (ii), we know by (20) that $U_g \quad g \;" \quad \ln(k_g + 1) > U_{g^0} \quad g^0 \;" \quad \ln(k_{g^0} + 1)$, and so $U_g \quad g \;" \quad \ln(k_g + 1) > U_{g^0} \quad g^0 \;" \quad \ln(k_{g^0} + m)$ for any $m \; 2 \; f \, 1; 2; : : : g$. 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 (20)–and hence a gainer–in the 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 $k_l \quad 1$ units for alternative l, we may apply (5) to calculate the possible gains in the objective function at the next iteration as

$U_l \quad l$ l, we may apply (5) to calculate the possible gains in the

l

The potential gains in (25) are all negative by (21); the potential gain in (26) is negative by (22). This means the optimality conditions have been met (and we are done), or else l is not the maximizer for (20) 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^{th} unit). For case (ii), we know by (21) that $U_l \; l \; " \; \ln(k_l) < U_{l'} \; l' \; " \; \ln(k_{l'})$ and so $U_l \; l \; " \; \ln(k_l) < U_{l'} \; l' \; " \; \ln(k_{l'} \; m)$ for any $m \in \{1; 2; \ldots\}$, $k_{l'} \; m \; 0$. Thus l could never be a maximizer in (20) with $(k_l \; 1)$ units of l once it has been a minimizer in (21). 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 swaps.

□

THEOREM 3

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 $U_{M+1} =$, where 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 $i = 1; 2; \ldots; M$ 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 go to alternative 1 because it solves (6). Observe that this must be the optimal 1-pack; for if the dummy alternative initially had =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 $i = 2; \ldots; M$. Observe that the alternative receiving the second unit is independent of n ($n \; 2$) and the resulting 2-pack must be the optimal 2-pack; for if the dummy alternative initially had =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

$U_i \quad \ln(k_i + 1)$, where the values k_i are the quantities already determined for the optimal (n)-pack at the previous iteration. Observe that only determines how many swaps are made, and the sequence of swaps is otherwise independent of . □

THEOREM 4.

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 -pack $(k_1; k_2; :::; k_M)$ 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 $U_i \quad \ln(k_i + 1) \quad U_j \quad \ln \; k_j$, which can be rearranged to yield the result.

□

THEOREM 5.

For $t = 1$, the index set $I_1(1; k_1; k_2; :::; k_M)$ reduces to a set of $m + 1$ vectors, each having dimension $(M + 1)$. One vector is $(1; 0; 0; :::)$, which captures the selection of the outside option; the remaining m vectors have a "1" in the position of the alternative included in the -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

have (using (27) with U_j replaced by $U_j + V_t(k_1; k_2; k_j \quad 1; :::; k_M)$ and U_0 replaced by $U_0 + V_t(k_1; k_2; :::; k_M)$):

$$V_{t+1}(k_1; k_2; :::; k_M) = \ln \left[\left(\sum_{\{j : k_j > 0\}} e^{U_j + V_t(k_1; k_2; k_j \quad 1; :::; k_M)} \right) + e^{U_0 + V_t(k_1; k_2; :::; k_M)} \right] + \quad ; \qquad (28)$$

which by the induction step

+1

= ln

and x_0 from the analysis. The index set reduces to

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

for $t \le n$ and the outside option terms in equations 28 and 29 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 $V_n(k_1;k_2;\dots;k_M) = \ln \left(\frac{n!}{k_1!k_2!\cdots k_M!} e^{\sum_{i=1}^{M} k_i U_i} \right) + n$, 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: $\max_{j:k_j>0} (U_j + \ _j + V_{n-1}(k_1;k_2;k_j - 1;\dots;k$

levels $x_i = q_i + 1$ and $x_j = 0; 1; \ldots; q_j - 1$). Because $q_j > k_j \geq k_i > q_i$, we must have $q_j \geq q_i + 2$ and so $q_j \prod_{l \neq j;i} (q_l + 1) > \prod_{l \neq j} (q_l + 1)$.

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 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

which implies

$$\frac{e^{U_j}}{q_j} \geq \frac{e^{U_i}}{q_i + 1};$$

and thus

$$\frac{e^{U_j}}{q_j - 1} \geq \frac{e^{U_i}}{q_i}$$
$$\frac{e^{U_j}}{q_j - 2} \geq \frac{e^{U_i}}{q_i - 1}$$
$$\vdots$$

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

$$\frac{e^{(z+1)U_j}}{\frac{q_j!}{(q_j - z - 1)!}} \geq \frac{e^{(z+1)U_i}}{\frac{(q_i+1)!}{(q_i - z)!}}:$$

This implies the bracketed term in the difference equation (30) 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. $a_i \geq 0$, $b_i \geq 0$, $i = 0; 1; \ldots; m$, $S^+ = \{i : a_i > b_i\}$, $S^- = \{i : a_i \leq b_i\}$, $\sum a_i \geq \sum b_i$; $\sum_{i \in S^+} \lambda_i a_i + \sum_{i \in S^-} \mu_i a_i \geq \sum_{i \in S^+} \lambda_i b_i + \sum_{i \in S^-} \mu_i b_i$, $0 \leq \mu_i \leq 1$, $\lambda_i = 1$

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 $\sum_{i \in S^+} \lambda_i(a_i - b_i) \geq \sum_{i \in S^-} \mu_i(a_i - b_i)$, which is a re-arrangement of the stated result.

□

THEOREM 7

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 $\gamma = \gamma + U_0$ 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 K-pack.

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

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

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

$$\begin{aligned}
\exp(V_{t-k}(K) + V_{t'+k}(K)) &\leq \exp(V_{t-k+1}(K) + V_{t'+k-1}(K)) \\
&\leq \exp(V_{t-k+2}(K) + V_{t'+k-2}(K)) \qquad (31) \\
&\vdots \\
&\leq \exp(V_t(K) + V_{t'}(K))
\end{aligned}$$

For any n-pack $K = (k_1, k_2, \ldots, k_M)$, consider any product having a [illegible] quantity. Suppose j is one such product. Define $\hat{K}_j = (k_1, \ldots, k_{j-1}, 0, k_{j+1}, \ldots, k_M)$. Observe that $\hat{K}_j$ must have $n - 1$ items or less. Because $U_0 = 0$, we may write the value function (see (14)) as

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

$$= \exp(t\lambda)\sum_{z=0}^{k_j}\frac{t!\exp(zU_j)}{z!(t-z)!}\exp\left[V_{t-z}(\hat{K}_j)\right]\exp((z-t)\lambda)$$

$$= \sum_{z=0}^{k_j}\binom{t}{z}\exp(z(U_j+\lambda))\exp\left[V_{t-z}(\hat{K}_j)\right].$$

It follows that

$$\exp[V_{t+1}(K)+V_{t-1}(K)] =$$

$$\sum_{z=0}^{k_j}\sum_{w=0}^{k_j}\binom{t+1}{z}\binom{t-1}{w}\exp((w+z)(U_j+\lambda))\exp\left[V_{t+1-z}(\hat{K}_j)+V_{t-1-w}(\hat{K}_j)\right] \quad (32)$$

and

$$\exp[V_t(K)+V_t(K)] = \sum_{z=0}^{k_j}\sum_{w=0}^{k_j}\binom{t}{z}\binom{t}{w}\exp((w+z)(U_j+\lambda))\exp\left[V_{t-z}(\hat{K}_j)+V_{t-w}(\hat{K}_j)\right]. \quad (33)$$

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 (32) and (33) as elements of a $(k_j+1)\times(k_j+1)$ matrix with rows $z = 0;1;2;\ldots;k_j$ and columns $w = 0;1;2;\ldots;k_j$, 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 (33) exceed those in equation (32). After dividing out the term $\exp(L(U_j+\lambda))$ in bo[illegible] (32) and (33), we must show

$$\sum_{l=0}^{L}\binom{t}{L-l}\binom{t}{l}\exp\left[V_{t-L+l}(\hat{K}_j)\right]\exp\left[V_{t-l}(\hat{K}_j)\right]$$

$$\geq \sum_{l=0}^{L}\binom{t+1}{L-l}\binom{t-1}{l}\exp\left[V_{t+1-L+l}(\hat{K}_j)\right]\exp\left[V_{t-1-l}(\hat{K}_j)\right]$$

term by its upper bound $\exp\left(-V_t(\hat{k}_j)\right) - \exp\left(-V_{t-L}(\hat{k}_j)\right)$ (by the induction hypothesis (31)) 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 (34). We will use a's to represent the combined coefficients for the left hand side of (34), and we will use b's to represent coefficients for the right hand side.

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

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

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

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 $\exp\left(V_{t-i}(\hat{k}_j)\right) - \exp\left(V_{t-L+i}(\hat{k}_j)\right)$ for $i = 0, 1, 2, \ldots, m$. Recall that the value $\exp\left(V_{t+1}(\hat{k}_j)\right) - \exp\left(V_{t-L-1}(\hat{k}_j)\right)$ in the right hand sum of (34) has been replaced with the (larger) term $\exp\left(V_t(\hat{k}_j)\right) - \exp\left(V_{t-L}(\hat{k}_j)\right)$ for the calculation of b_0. This means

$$
\sum_{i=0}^{m} a_i \left[\exp\left(V_{t-i}(\hat{k}_j)\right) - \exp\left(V_{t-L+i}(\hat{k}_j)\right)\right] =
$$

$$
\sum_{l=0}^{L} \begin{pmatrix} t \\ L-l \end{pmatrix} \begin{pmatrix} t \\ l \end{pmatrix} \left[\exp\left(V_{t-L+l}(\hat{k}_j)\right) - \exp\left(V_{t-l}(\hat{k}_j)\right)\right] \tag{35}
$$

but

$$
\sum_{i=0}^{m} b_i \left[\exp\left(V_{t-i}(\hat{k}_j)\right) - \exp\left(V_{t-L+i}(\hat{k}_j)\right)\right] \geq
$$

$$
\sum_{l=0}^{L} \binom{t+1}{L-l} \binom{t-1}{l} \left[\exp\left(V_{t+1-L+l}(\hat{k}_j)\right) - \exp\left(V_{t-1-l}(\hat{k}_j)\right)\right] \tag{36}
$$

We now show

$$
\sum_{i=0}^{m} a_i \left[\exp\left(V_{t-i}(\hat{k}_j)\right) - \exp\left(V_{t-L+i}(\hat{k}_j)\right)\right] \geq \sum_{i=0}^{m} b_i \left[\exp\left(V_{t-i}(\hat{k}_j)\right) - \exp\left(V_{t-L+i}(\hat{k}_j)\right)\right] \tag{37}
$$

It is straightforward to show that $a_0 \geq b_0$. It is also straightforward to show $a_m > b_m$. For

$i = 1, \ldots, m^* - 1$, we have

$$a_i - b_i = \binom{t}{L-i}\binom{t}{i}\left(2 - \frac{t+1}{t}\left(\frac{i}{L+1-i} + \frac{L-i}{i+1}\right)\right). \tag{38}$$

The function $g(x) = \frac{x}{L+1-x} + \frac{L-x}{x+1}$ appearing in the bracketed term of (38) is strictly decreasing for $0 \le x \le m-1$. For any given t, the function $2 - \frac{t+1}{t} g(x)$ is therefore strictly increasing in . Consequently, there is an index $0 < m^* \le m$ such that $a_i - b_i > 0$ for $i \ge m^*$ and $a_i - b_i \le 0$ for $i < m^*$. Define

$$\alpha_i = \frac{\exp\left(V_{t-i}(\hat{k}_j)\right) - \exp\left(V_{t-L+i}(\hat{k}_j)\right)}{\exp\left(V_{t-m^*}(\hat{k}_j)\right) - \exp\left(V_{t-L+m^*}(\hat{k}_j)\right)} \quad i = 0, 1, \ldots, m^* - 1$$

and

$$\beta_i = \frac{\exp\left(V_{t-i}(\hat{k}_j)\right) - \exp\left(V_{t-L+i}(\hat{k}_j)\right)}{\exp\left(V_{t-m^*}(\hat{k}_j)\right) - \exp\left(V_{t-L+m^*}(\hat{k}_j)\right)} \quad i = m^*, m^*+1, \ldots, m.$$

$P_{i=m}^{m} \; {}_i b_i$. Multiplying both sides of this inequality by the term $\exp\left(V_{t\,m}(\hat{k}_j)\right) \exp\left(V_{t\;L+m}(\hat{k}_j)\right)$ yields the inequality (37) and the theorem is proved.

□

Appendix B (Numerical Experiments)

We assume there are 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 $U_1 \quad U_2 \qquad U_n$. The indices for the consumer ( ) and the consumption time ( ) are suppressed. The cdf for alternative 's error ${}_i$ is F_i and its density is f_i.

The probability that the consumer chooses item i is

$$P(i) = \mathrm{Prob}\{U_i + {}_i \quad U_j + "_j \quad \forall j;\, j \neq i\};$$

which implies the marginal choice probability

$$P(i) = \int_{1}^{1} \prod_{j \neq i} F_j\,(U_i \quad U_j + \;)\, f_i(\;)\, d\; : \tag{40}$$

We will have frequent occasion to calculate the expected value of the maximum of random variables having the general form $X_i = c_i + {}_i$ for $i \in A \quad \{1; 2; 3 ::: ; n\}$. The cdf $G_A(t)$ for max $::: ; n\}$

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 nested loops organized in a particular fashion. The outermost loop corresponds to the most attractive 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 $k_1 = 0; 1; 2 : : : ; n$; the next loop uses $k_2 = 0; 1; 2 : : : n \quad k_1$; the third loop uses $k_3 = 0; 1; 2 : : : n \quad k_1 \quad k_2$ and so on. The innermost loop uses $k_n = 0; 1; 2 : : : n \quad k_1 \quad k_2 \qquad k_{n \ 1}$. Calculating the value function for each -pack ($j \quad n$) 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

Table 1: Computed tility Parameters for 3-Packs using niform ( ) and Normal (N)

Scenario	U_1( )	U_2( )	U_3( )	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

pothetical case (33.33%, 33.33%, 33.33%) used in x3:4.) 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 $U_i = \ln(!_i)$ where $!_i$ 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 (40) and a simple weighted least squares optimization model where the $!_i$ served as targets. All values are unique up to an additive shift. The 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 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 $f(i) = \frac{\prod_{l=1}^{i} X_l}{\sum_{i=1}^{6} \prod_{l=1}^{i} X_l}$, which ensured $f(1)$ $f(2)$ $f(6)$ 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. pon 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, (i).0049, and .0000007.on examini9.728el aslichoice frequency belocee000007.27(boice)-9h4frecee000007.27

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 -packs implicitly ignores such packs, we would expect these distortions, when they occur, to have no meaningful impact on our results.