STRUCTURING THE NEW PRODUCT DEVELOPMENT PIPELINE-- APPENDICES

Min Ding and Jehoshua Eliashberg

APPENDIX A: Proof of Lemma 1, Propositions 2-3, and Impact of Correlation

Stage k Expected Profit Formulation

The expression for the expected profit may be formally defined and proved in conjunction with Lemma 1, and should technically follow Lemma 1. We define it earlier in the text to conform to the structure. The proposed expression for the expected profit at stage k is based on the assumption that the expected profit at stage k-1 is concave prior to reaching a unique optimal value (n_{k-1}*), however, the proof for Lemma 1 (see below) does not rely on this expression *a prior*. Essentially, given stage 1 results (strictly concave), the expression for the expected profit at stage 2 is valid and we could thus use it to prove the 1st step in Lemma 1 (i.e., strictly concave for stage 2). In the 2nd step, we already assume the expected profit for stage j is strictly concave, thus the expression for expected profit at stage j+1 is valid as well, and we could use it to show the expected profit is strictly concave for stage j+1. Thus prove both the validity of the expression and the strict concavity of the expected profit.

Lemma 1: $E[\pi_k(n_k)]$ is a strictly concave function with a unique global maximum at n_k*, where n_k* is implicitly defined by the following equation:

$$p_k \sum_{s_k=0}^{n_k^*-1} \left\{ \Pr(s_k \mid p_k, n_k^*) \left[E[\pi_{k-1}(s_k+1)] - E[\pi_{k-1}(s_k)] \right] \right\} - c_k = 0 \qquad \text{(A1)}$$

& (12)

k is a positive integer and k≥2.

Proof:

This proof uses induction and is comprised of two steps:

Step 1: show the proposition holds when k=2;

Step 2: show the proposition holds for k=j+1, if we assume the proposition holds for k=j

(j is a positive integer and j≥2);

In order to facilitate the proof, we need to first reformulate the expression of expected return at stage k. For convenience, we will use following abbreviation:

$\Pr(s_k|n_k) = \Pr(s_k|p_k, n_k)$

The outcome of m+1 parallel approaches at stage k could be viewed as a combination of outcomes from two groups. One group contains one approach, the other contains m approaches. Thus, the expected return at stage k for m+1 approaches could be reformulated as following:

$$
\begin{aligned}
E[\pi_k(m+1)] &= \sum_{s_k=n_{k-1}^*}^{m+1} \left[\Pr(s_k|m+1)E[\pi_{k-1}(n_{k-1}^*)]\right] + \sum_{s_k=0}^{n_{k-1}^*-1} \left[\Pr(s_k|m+1)E[\pi_{k-1}(s_k)]\right] - (m+1)c_k \\
&= p_k \left\{ \sum_{s_k=n_{k-1}^*-1}^{m} \left[\Pr(s_k|m)E[\pi_{k-1}(n_{k-1}^*)]\right] + \sum_{s_k=0}^{n_{k-1}^*-2} \left[\Pr(s_k|m)E[\pi_{k-1}(s_k+1)]\right] \right\} \\
&\quad + (1-p_k)\left\{ \sum_{s_k=n_{k-1}^*}^{m} \left[\Pr(s_k|m)E[\pi_{k-1}(n_{k-1}^*)]\right] + \sum_{s_k=0}^{n_{k-1}^*-1} \left[\Pr(s_k|m)E[\pi_{k-1}(s_k)]\right] \right\} - (m+1)c_k
\end{aligned}
\qquad \text{(A2)}
$$

As a result, we could now simplify the incremental expected return (the discrete equivalent of 1st order derivative against m) offered by the (m+1)th approach:

$$
\begin{aligned}
E[\pi_k(m+1)] - E[\pi_k(m)] &= p_k \left\{ \sum_{s_k=n_{k-1}^*-1}^{m} \left[\Pr(s_k|m)E[\pi_{k-1}(n_{k-1}^*)]\right] + \sum_{s_k=0}^{n_{k-1}^*-2} \left[\Pr(s_k|m)E[\pi_{k-1}(s_k+1)]\right] \right\} \\
&\quad + (1-p_k)\left\{ \sum_{s_k=n_{k-1}^*}^{m} \left[\Pr(s_k|m)E[\pi_{k-1}(n_{k-1}^*)]\right] + \sum_{s_k=0}^{n_{k-1}^*-1} \left[\Pr(s_k|m)E[\pi_{k-1}(s_k)]\right] \right\} - (m+1)c_k \\
&\quad - \sum_{s_k=n_{k-1}^*}^{m} \left[\Pr(s_k|m)E[\pi_{k-1}(n_{k-1}^*)]\right] - \sum_{s_k=0}^{n_{k-1}^*-1} \left[\Pr(s_k|m)E[\pi_{k-1}(s_k)]\right] + mc_k \\
&= p_k \left\{ \sum_{s_k=n_{k-1}^*-1}^{m} \left[\Pr(s_k|m)E[\pi_{k-1}(n_{k-1}^*)]\right] - \sum_{s_k=n_{k-1}^*}^{m} \left[\Pr(s_k|m)E[\pi_{k-1}(n_{k-1}^*)]\right] \right\} \\
&\quad + p_k \left\{ \sum_{s_k=0}^{n_{k-1}^*-2} \left[\Pr(s_k|m)E[\pi_{k-1}(s_k+1)]\right] - \sum_{s_k=0}^{n_{k-1}^*-1} \left[\Pr(s_k|m)E[\pi_{k-1}(s_k)]\right] \right\} - c_k \\
&= p_k \Pr(n_{k-1}^*-1|m)E[\pi_{k-1}(n_{k-1}^*)] - p_k \Pr(n_{k-1}^*-1|m)E[\pi_{k-1}(n_{k-1}^*-1)] \\
&\quad + p_k \sum_{s_k=0}^{n_{k-1}^*-2} \left[\Pr(s_k|m)\left(E[\pi_{k-1}(s_k+1)] - E[\pi_{k-1}(s_k)]\right)\right] - c_k \\
&= p_k \sum_{s_k=0}^{n_{k-1}^*-1} \left[\Pr(s_k|m)\left(E[\pi_{k-1}(s_k+1)] - E[\pi_{k-1}(s_k)]\right)\right] - c_k
\end{aligned}
$$

(A3)

The discrete equivalent of the 2nd order derivative against m could be obtained by calculating the difference between adjacent incremental expected return (1st order derivative):

$$\left[E[\pi_k(m+2)]-E[\pi_k(m+1)]\right]-\left[E[\pi_k(m+1)]-E[\pi_k(m)]\right] \tag{A4}$$

Using (A3) and similar partition technique, we could obtain:

$$\begin{aligned}
\left[E[\pi_k(m+2)]-E[\pi_k(m+1)]\right] &= p_k \sum_{s_k=0}^{n^*_{k-1}-1}\left[\Pr(s_k|m+1)\left(E[\pi_{k-1}(s_k+1)]-E[\pi_{k-1}(s_k)]\right)\right]-c_k \\
&= p_k\left\{p_k \sum_{s_k=0}^{n^*_{k-1}-2}\left[\Pr(s_k|m)\left(E[\pi_{k-1}(s_k+2)]-E[\pi_{k-1}(s_k+1)]\right)\right]+(1-p_k)\sum_{s_k=0}^{n^*_{k-1}-1}\left[\Pr(s_k|m)\left(E[\pi_{k-1}(s_k+1)]-E[\pi_{k-1}(s_k)]\right)\right]\right\}-c_k \\
&= p_k^2\left\{\sum_{s_k=0}^{n^*_{k-1}-2}\left[\Pr(s_k|m)\left(E[\pi_{k-1}(s_k+2)]-E[\pi_{k-1}(s_k+1)]\right)\right]-\sum_{s_k=0}^{n^*_{k-1}-1}\left[\Pr(s_k|m)\left(E[\pi_{k-1}(s_k+1)]-E[\pi_{k-1}(s_k)]\right)\right]\right\} \\
&\quad + p_k\sum_{s_k=0}^{n^*_{k-1}-1}\left[\Pr(s_k|m)\left(E[\pi_{k-1}(s_k+1)]-E[\pi_{k-1}(s_k)]\right)\right]-c_k \\
&= p_k^2\left\{\sum_{s_k=0}^{n^*_{k-1}-2}\left[\Pr(s_k|m)\left(E[\pi_{k-1}(s_k+2)]-E[\pi_{k-1}(s_k+1)]\right)\right]-\sum_{s_k=0}^{n^*_{k-1}-1}\left[\Pr(s_k|m)\left(E[\pi_{k-1}(s_k+1)]-E[\pi_{k-1}(s_k)]\right)\right]\right\} \\
&\quad + \left[E[\pi_k(m+1)]-E[\pi_k(m)]\right]
\end{aligned} \tag{A5}$$

Thus, we could obtain the equivalent 2nd derivative by rearranging (A5):

$$\begin{aligned}
&\left[E[\pi_k(m+2)]-E[\pi_k(m+1)]\right]-\left[E[\pi_k(m+1)]-E[\pi_k(m)]\right] \\
&= p_k^2\left\{\sum_{s_k=0}^{n^*_{k-1}-2}\left[\Pr(s_k|m)\left(E[\pi_{k-1}(s_k+2)]-E[\pi_{k-1}(s_k+1)]\right)\right]-\sum_{s_k=0}^{n^*_{k-1}-1}\left[\Pr(s_k|m)\left(E[\pi_{k-1}(s_k+1)]-E[\pi_{k-1}(s_k)]\right)\right]\right\} \\
&= p_k^2\left\{\sum_{s_k=0}^{n^*_{k-1}-2}\left[\Pr(s_k|m)\left[\left(E[\pi_{k-1}(s_k+2)]-E[\pi_{k-1}(s_k+1)]\right)-\left(E[\pi_{k-1}(s_k+1)]-E[\pi_{k-1}(s_k)]\right)\right]\right]-\left[\Pr(n^*_{k-1}-1|m)\left(E[\pi_{k-1}(n^*_{k-1})]-E[\pi_{k-1}(n^*_{k-1}-1)]\right)\right]\right\}
\end{aligned} \tag{A6}$$

Now we could prove (A1) holds for all k (k is a positive integer and k≥2) using induction.

Step 1: k=2.

(A6) could be simplified to:

$$\begin{aligned}
&\left[E[\pi_2(m+2)]-E[\pi_2(m+1)]\right]-\left[E[\pi_2(m+1)]-E[\pi_2(m)]\right] \\
&= p_2^2\left\{\sum_{s_2=0}^{n^*_1-2}\left[\Pr(s_2|m)\left[\left(E[\pi_1(s_2+2)]-E[\pi_1(s_2+1)]\right)-\left(E[\pi_1(s_2+1)]-E[\pi_1(s_2)]\right)\right]\right]\right\} \\
&\quad - p_2^2\left[\Pr(n^*_1-1|m)\left(E[\pi_1(n^*_1)]-E[\pi_1(n^*_1-1)]\right)\right]
\end{aligned} \tag{A7}$$

Based results from stage 1, we could obtain:

$$\begin{aligned}
E[\pi_1(s_2+1)]-E[\pi_1(s_2)] &= [1-(1-p_1)^{s_2+1}]E[\pi_0(s_1>0)]-(s_2+1)c_1-[1-(1-p_1)^{s_2}]E[\pi_0(s_1>0)]+s_2c_1 \\
&= p_1(1-p_1)^{s_2}E[\pi_0(s_1>0)]-c_1
\end{aligned} \tag{A8}$$

thus, assuming p_1 does not equal 0 nor 1,

$$\begin{aligned}&\left[E[\pi_1(s_2+2)]-E[\pi_1(s_2+1)]\right]-\left[E[\pi_1(s_2+1)]-E[\pi_1(s_2)]\right]\\&=\left[p_1(1-p_1)^{s_2+1}E[\pi_0(s_1>0)]-c_1\right]-\left[p_1(1-p_1)^{s_2}E[\pi_0(s_1>0)]-c_1\right]\\&=-p_1^{\,2}(1-p_1)^{s_2}E[\pi_0(s_1>0)]<0\end{aligned}\tag{A9}$$

Given (A9), the first term (summation) of (A7) must be smaller than 0. Given the definition of n_1*, the second term $-p_2^{\,2}\left[\Pr(n_1^*-1|m)\left(E\pi_1(n_1^*)-E\pi_1(n_1^*-1)\right)\right]$ also must be equal or smaller than 0.

As a result, (A7) must be smaller than 0, i.e.,

$$\left[E[\pi_2(m+2)]-E[\pi_2(m+1)]\right]-\left[E[\pi_2(m+1)]-E[\pi_2(m)]\right]<0\tag{A10}$$

Thus, we have proved that the expected return at stage 2 is strictly concave. The value n_2*, obtained by setting the 1st order derivative (A3) equals to 0, is the unique value that maximizes the expected return at stage 2.

Step 2: Assuming the proposition holds for k=j (j is a positive integer and j≥2), thus

$$\left[E[\pi_j(m+2)]-E[\pi_j(m+1)]\right]-\left[E[\pi_j(m+1)]-E[\pi_j(m)]\right]<0\tag{A11}$$

for stage j+1, the 2nd derivative could be represented by

$$\begin{aligned}&\left[E[\pi_{j+1}(m+2)]-E[\pi_{j+1}(m+1)]\right]-\left[E[\pi_{j+1}(m+1)]-E[\pi_{j+1}(m)]\right]\\&=p_{j+1}^{\,2}\left\{\sum_{s_{j+1}=0}^{n_j^*-2}\left[\Pr(s_{j=1}|m)\left[\left(E[\pi_j(s_{j+1}+2)]-E[\pi_j(s_{j+1}+1)]\right)-\left(E[\pi_j(s_{j+1}+1)]-E[\pi_j(s_{j+1})]\right)\right]\right]\right\}\\&-p_{j+1}^{\,2}\left[\Pr(n_j^*-1|m)\left(E[\pi_j(n_j^*)]-E[\pi_j(n_j^*-1)]\right)\right]\end{aligned}\tag{A12}$$

Similar to step 1, (A12) is smaller than 0 because first term in (A12) is smaller than 0 because of (A11), and second term is smaller than or equal to 0 by definition of n_j*. Thus, the expected return function at stage j+1 is also strictly concave and thus proposition (A1) holds.

Combine the results from step 1 and step 2, we conclude the proposition holds for all k, where k is a positive integer and k≥2.

Q.E.D.

Proposition 2:

For stage k (k≥2):

n_k* increases when c_k decreases;

n_k* reaches maximum at an interior value of p_k(between 0 and 1);

An approximation (upper bound) for n_k* is given by

$$n_k^* < \frac{\ln c_k - \ln(p_1 p_2 \cdots p_k E[\pi_0(s_1 > 0)])}{\ln(1 - p_1 p_2 \cdots p_k)} \qquad \text{(A13) \& (13)}$$

Proof:

When the optimal number of approaches (n_k*) is funded, the marginal benefit of the n_k*th approach is equal to the marginal cost (c_k). If c_k decreases, the marginal benefit of the n_k*th approach will then be larger than its marginal cost. Given the concavity (from Lemma 1), then the new optimal number must be larger than the original optimal number (i.e., n_k* will increase).

The observation that n_k* reaches maximum at an interior value of p_k(between 0 and 1) could be made based on the fact that there is no marginal benefit when p_k equals either 0 or 1.

In order to prove the upper bound, we will first use induction to prove that

$$E\pi_k[(n_k + 1)] - E[\pi_k(n_k)] < p_1 p_2 \cdots p_k E[\pi_0](1 - p_1 p_2 \cdots p_k)^{n_k} - c_k \qquad \text{(A14)}$$

For k=2:

$$E[\pi_2(n_2+1)]-E[\pi_2(n_k)] = p_2\sum_{s_2=0}^{n_1^*-1}\left\{\Pr(s_2\mid p_2,n_2)[E[\pi_1(s_2+1)]-E[\pi_1(s_2)]]\right\}-c_2$$

$$= p_2\sum_{s_2=0}^{n_1^*-1}\left\{\Pr(s_2\mid p_2,n_2)[p_1(1-p_1)^{s_2}E[\pi_0]-c_1]\right\}-c_2$$

$$= p_1p_2E[\pi_0]\sum_{s_2=0}^{n_1^*-1}\left\{\binom{n_2}{s_2}p_2^{s_2}(1-p_2)^{n_2-s_2}(1-p_1)^{s_2}\right\}-p_2c_1\sum_{s_2=0}^{n_1^*-1}\left\{\Pr(s_2\mid p_2,n_2)\right\}-c_2$$

$$= p_1p_2E[\pi_0]\sum_{s_2=0}^{n_1^*-1}\left\{\binom{n_2}{s_2}\left(\frac{p_2(1-p_1)}{1-p_2}\right)^{s_2}(1-p_2)^{n_2}\right\}-p_2c_1\sum_{s_2=0}^{n_1^*-1}\left\{\Pr(s_2\mid p_2,n_2)\right\}-c_2$$

$$= p_1p_2E[\pi_0]\frac{(1-p_2)^{n_2}}{\left(\frac{1-p_2}{1-p_1p_2}\right)^{n_2}}\sum_{s_2=0}^{n_1^*-1}\left\{\binom{n_2}{s_2}\left(\frac{p_2(1-p_1)}{1-p_2}\right)^{s_2}\left(\frac{1-p_2}{1-p_1p_2}\right)^{n_2}\right\}-p_2c_1\sum_{s_2=0}^{n_1^*-1}\left\{\Pr(s_2\mid p_2,n_2)\right\}-c_2$$

$$= p_1p_2E[\pi_0](1-p_1p_2)^{n_2}\sum_{s_2=0}^{n_1^*-1}\left\{\Pr(s_2\mid \frac{p_2(1-p_1)}{1-p_1p_2},n_2)\right\}-p_2c_1\sum_{s_2=0}^{n_1^*-1}\left\{\Pr(s_2\mid p_2,n_2)\right\}-c_2$$

$$< p_1p_2E[\pi_0](1-p_1p_2)^{n_2}\sum_{s_2=0}^{n_1^*-1}\left\{\Pr(s_2\mid \frac{p_2(1-p_1)}{1-p_1p_2},n_2)\right\}-c_2$$

$$< p_1p_2E[\pi_0](1-p_1p_2)^{n_2}-c_2 \qquad \text{(A15)}$$

Assume the inequality for k=m-1 holds, i.e.,

$$E[\pi_{m-1}(n_{m-1}+1)]-E[\pi_{m-1}(n_{m-1})] < p_1p_2\cdots p_{m-1}E[\pi_0](1-p_1p_2\cdots p_{m-1})^{n_{m-1}}-c_{m-1} \qquad \text{(A16)}$$

then for k=m,

$$E[\pi_m(n_m+1)]-E[\pi_m(n_m)] = p_m \sum_{s_m=0}^{n_{m-1}^*-1}\left\{\Pr(s_m \mid p_m,n_m)[E[\pi_{m-1}(s_m+1)]-E[\pi_{m-1}(s_m)]]\right\}-c_m$$

$$< p_m \sum_{s_m=0}^{n_{m-1}^*-1}\left\{\Pr(s_m \mid p_m,n_m)[p_1p_2\cdots p_{m-1}E[\pi_0](1-p_1p_2\cdots p_{m-1})^{s_m}-c_{m-1}]\right\}-c_m$$

$$= p_1p_2\cdots p_m E[\pi_0]\sum_{s_m=0}^{n_{m-1}^*-1}\left\{\binom{n_m}{s_m} p_m^{s_m}(1-p_m)^{n_m-s_m}(1-p_1p_2\cdots p_{m-1})^{s_m}\right\}-p_mc_{m-1}\sum_{s_m=0}^{n_{m-1}^*-1}\left\{\Pr(s_m \mid p_m,n_m)\right\}-c_m$$

$$= p_1p_2\cdots p_m E[\pi_0]\sum_{s_m=0}^{n_{m-1}^*-1}\left\{\binom{n_m}{s_m}\left(\frac{p_m(1-p_1p_2\cdots p_{m-1})}{1-p_m}\right)^{s_m}(1-p_m)^{n_m}\right\}-p_mc_{m-1}\sum_{s_m=0}^{n_{m-1}^*-1}\left\{\Pr(s_m \mid p_m,n_m)\right\}-c_m$$

$$= p_1p_2\cdots p_m E[\pi_0]\frac{(1-p_m)^{n_m}}{\left(\dfrac{1-p_m}{1-p_1p_2\cdots p_m}\right)^{n_m}}\sum_{s_m=0}^{n_{m-1}^*-1}\left\{\binom{n_m}{s_m}\left(\frac{p_m(1-p_1p_2\cdots p_{m-1})}{1-p_m}\right)^{s_m}\left(\frac{1-p_m}{1-p_1p_2\cdots p_m}\right)^{n_m}\right\}$$

$$-p_mc_{m-1}\sum_{s_m=0}^{n_{m-1}^*-1}\left\{\Pr(s_m \mid p_m,n_m)\right\}-c_m$$

$$= p_1p_2\cdots p_m E[\pi_0]\frac{(1-p_m)^{n_m}}{\left(\dfrac{1-p_m}{1-p_1p_2\cdots p_m}\right)^{n_m}}\sum_{s_m=0}^{n_{m-1}^*-1}\left\{\Pr\left(s_2 \mid \frac{p_m(1-p_1p_2\cdots p_{m-1})}{1-p_1p_2\cdots p_m},n_m\right)\right\}-p_mc_{m-1}\sum_{s_m=0}^{n_{m-1}^*-1}\left\{\Pr(s_m \mid p_m,n_m)\right\}-c_m$$

$$< p_1p_2\cdots p_m E[\pi_0](1-p_1p_2\cdots p_m)^{n_m}\sum_{s_m=0}^{n_{m-1}^*-1}\left\{\Pr\left(s_2 \mid \frac{p_m(1-p_1p_2\cdots p_{m-1})}{1-p_1p_2\cdots p_m},n_m\right)\right\}-c_m$$

$$< p_1p_2\cdots p_m E[\pi_0](1-p_1p_2\cdots p_m)^{n_m}-c_m \tag{A17}$$

Thus, the inequality (A14) holds for all k, where k≥2.

As a result, when $n_k = n_k^*$,

$$E[\pi_k(n_k^*+1)]-E[\pi_k(n_k^*)] = 0 < p_1p_2\cdots p_k E[\pi_0](1-p_1p_2\cdots p_k)^{n_k^*}-c_k$$

$$\Rightarrow (1-p_1p_2\cdots p_k)^{n_k^*} > \frac{c_k}{p_1p_2\cdots p_k E[\pi_0]}$$

$$\Rightarrow n_k^*\ln(1-p_1p_2\cdots p_k) > \ln\frac{c_k}{p_1p_2\cdots p_k E[\pi_0]}$$

$$\Rightarrow n_k^* < \frac{\ln c_k - \ln(p_1p_2\cdots p_k E[\pi_0])}{\ln(1-p_1p_2\cdots p_k)}$$

Q.E.D.

Proposition 3:

If, for all k (k≥2),

$$c_k < p_k \sum_{s_k=0}^{n_{k-1}^*-1} \{\Pr(s_k \mid p_k, n_{k-1}^*)[E[\pi_{k-1}(p_{k-1}, s_k+1, c_{k-1}, E\pi_{k-2})] - E[\pi_{k-1}(p_{k-1}, s_k, c_{k-1}, E\pi_{k-2})]]\} = \bar{c}_k$$

(A18) & (14)

then the NPD pipeline will take the shape of a **funnel** ($n_k^* > n_{k-1}^*$). Otherwise the pipeline will be a **tunnel** shape ($n_k^* = n_{k-1}^*$).

Proof:

From Lemma 1, we know that

$$p_k \sum_{s_k=0}^{n_{k-1}^*-1} \{\Pr(s_k \mid p_k, n_{k-1}^*)[E[\pi_{k-1}(p_{k-1}, s_k+1, c_{k-1}, E\pi_{k-2})] - E[\pi_{k-1}(p_{k-1}, s_k, c_{k-1}, E\pi_{k-2})]]\} = \bar{c}_k$$

represents the marginal benefit of supporting the n_{k-1}^*th approach in stage k, in other words, supporting the same number of approaches in both stage k-1 and k.

If (A18) holds, then the marginal benefit of supporting this approach is larger than the marginal cost (c_k), and thus this approach should be supported. Given the concavity (from Lemma 1), n_k^* must be larger than n_{k-1}^* and the pipeline will therefore have a funnel structure.

If (A18) does not hold, then the marginal benefit of supporting this approach is smaller than the marginal cost (c_k), and this approach should either not be supported or be the last one to be supported (if equality holds). Given the concavity (from Lemma 1), n_k^* must be the same or smaller than n_{k-1}^*. However, given the constraint that approaches have to pass earlier stages to move to later stages, the resulting pipeline will have a tunnel structure with its width equals to n_k^*.

Q.E.D.

Impact of Correlated Approaches:

The optimal number of approaches at Stage 1 will increase as the (positive) correlation becomes weaker.

Proof:

Let's assume that the success of an individual approach depends on two factors, one is a common obstacle presented in all approaches (with probability of success p_c), the other is the factor unique to each approach (with probability of success p_1/p_c). This reduces to the base model if $p_c=1$.

The pair-wise correlation could be calculated as:

$$corr = \frac{\operatorname{cov}(Y_1,Y_2)}{\sigma_1\sigma_2} = \frac{E(Y_1Y_2)-E(Y_1)E(Y_2)}{p_1(1-p_1)} = \frac{p_c({p_1}/{p_c})^2 - p_1^2}{p_1(1-p_1)} = \frac{p_1}{1-p_1}\left(\frac{1}{p_c}-1\right) \qquad \text{(A19)}$$

Thus, smaller p_c (most of the uncertainty is due to the common obstacle) will result in higher correlation.

Under this scenario, the expected return at stage 1 is:

$$E[\pi_1(n_1)] = p_c[1-(1-\frac{p_1}{p_c})^{n_1}]E[\pi_0(s_1>0)] - n_1c_1 \qquad \text{(A20)}$$

So, the marginal change in return is

$$E[\pi_1(n_1+1)] - E[\pi_1(n_1)] = p_c[\frac{p_1}{p_c}(1-\frac{p_1}{p_c})^{n_1}]E[\pi_0(s_1>0)] - c_1 = p_1(1-\frac{p_1}{p_c})^{n_1}E[\pi_0(s_1>0)] - c_1 \qquad \text{(A21)}$$

the above quantity becomes 0 when $n_1 = n_1^*$. In which case, we could solve for n_1^*:

$$n_1^* = \frac{\ln c_1 - \ln\{p_1E[\pi_0(s_1>0)]\}}{\ln(1-\frac{p_1}{p_c})} \qquad \text{(A22)}$$

we could obtain the 1st derivative:

$$\frac{\partial n_1^*}{\partial p_c} = \frac{\ln c_1 - \ln(p_1E[\pi_0(s_1>0)])}{\left(\ln(1-\frac{p_1}{p_c})\right)^2}\left(-\frac{\frac{p_1}{p_c^2}}{1-\frac{p_1}{p_c}}\right) > 0 \qquad \text{(A23)}$$

Thus, the optimal number increases as the probability of success for the common factor increases. Equivalently, the optimal number increases as the correlation decreases.

Q.E.D.

APPENDIX B: GEOMETRIC ANALYSIS OF ONE STAGE AND TWO STAGE SCENARIOS

The purpose of this analysis is to shed light to the optimal pipeline structure and be used as managerial guidelines when the specific parameter values are not available, or could not be accurately (or cost-effectively) obtained.

One-stage scenario:

First, we examine the condition under which at least one approach should be funded:

$$p_1 * E[\pi_0(s_1 > 0)] - c_1 > 0 \Leftrightarrow c_1 < p_1 E[\pi_0(s_1 > 0)] \tag{B1}$$

Given $E[\pi_1(n_1)]$ is concave over n_1 (Proposition 1), the condition under which two or more approaches should be funded simultaneously is following:

$$\begin{aligned} & E[\pi_1(2)] > E[\pi_1(1)] \\ & \Leftrightarrow \left[-(1-p_1)^2\right] E[\pi_0(s_1 > 0)] - 2c_1 > p_1 E[\pi_0(s_1 > 0)] - c_1 \\ & \Leftrightarrow c_1 < p_1(1-p_1)E[\pi_0(s_1 > 0)] \end{aligned} \tag{B2}$$

In general, the condition under which funding n_1 approaches is better than funding n_1-1 approaches could be expressed as:

$$\begin{aligned} & E[\pi_1(n_1)] > E[\pi_1(n_1 - 1)] \\ & \Leftrightarrow \left[-(1-p_1)^{n_1}\right] E[\pi_0(s_1 > 0)] - n_1 c_1 > \left[-(1-p_1)^{n_1-1}\right] E[\pi_0(s_1 > 0)] - (n_1 - 1)c_1 \\ & \Leftrightarrow c_1 < p_1(1-p_1)^{n_1-1} E[\pi_0(s_1 > 0)] \end{aligned} \tag{B3}$$

The relationship between the critical c_1 and the probability of success (p_1) is following:

$$\frac{\partial c_1}{\partial p_1} = (1-p_1)^{n_1-2} E[\pi_0(s_1 > 0)](1 - p_1 n_1) \begin{cases} < 0 & p_1 > 1/n_1 \\ > 0 & p_1 < 1/n_1 \end{cases} \tag{B4}$$

$$\frac{\partial^2 c_1}{\partial p_1^2} = (1-p_1)^{n_1-3} E[\pi_0(s_1 > 0)](n_1 - 1)(p_1 n_1 - 2) \begin{cases} < 0 & p_1 < 2/n_1 \\ > 0 & p_1 > 2/n_1 \end{cases} \tag{B5}$$

Based on the above analysis, the optimal strategy under various conditions could be represented in Figure 4.

<u>Two-stage scenario:</u>

The conditions for the last NPD stage could be obtained from the earlier analysis for one-stage process. Assuming $p = p_1 * p_2 \quad and \quad c = c_1 + c_2$, the conditions for the first NPD stage could all be expressed as inequalities between c_1 and a function of the p_1, p, c, and $E[\pi_0(s_1 > 0)]$:

Given $n_1^* = 1$,

$$E[\pi_2(1)] > E[\pi_2(0)]$$
$$\Leftrightarrow p_2(p_1 E[\pi_0(s_1 > 0)] - c_1) - c_2 > 0 \Leftrightarrow c_1 < \frac{p_1}{p_1 - p}(c - pE[\pi_0(s_1 > 0)]) \tag{B6}$$

$$E[\pi_2(2)] > E[\pi_2(1)]$$
$$\Leftrightarrow \left[p_2^2 + 2p_2(1-p_2)\right]\left\{p_1 E[\pi_0(s_1 > 0)] - c_1\right\} - 2c_2 > p_2(p_1 E[\pi_0(s_1 > 0)] - c_1) - c_2$$
$$\Leftrightarrow p_2(1-p_2)(p_1 E[\pi_0(s_1 > 0)] - c_1) - c_2 > 0 \tag{B7}$$
$$\Leftrightarrow c_1 > \frac{(c - pE[\pi_0(s_1 > 0)])p_1^2 + p^2 E[\pi_0(s_1 > 0)]p_1}{p_1^2 - pp_1 + p^2}$$

Given $n_1^* > 1$,

$$E[\pi_2(1)] > E[\pi_2(0)]$$
$$\Leftrightarrow p_2(p_1 E[\pi_0(s_1 > 0)] - c_1) - c_2 > 0 \Leftrightarrow c_1 < \frac{p_1}{p_1 - p}(c - pE[\pi_0(s_1 > 0)]) \tag{B8}$$

$$E[\pi_2(2)] > E[\pi_2(1)]$$
$$\Leftrightarrow p_2^2\left\{\left[-(1-p_1)^2\right]E[\pi_0(s_1 > 0)] - 2c_1\right\} + 2p_2(1-p_2)(p_1 E[\pi_0(s_1 > 0)] - c_1) - 2c_2 > p_2(p_1 E[\pi_0(s_1 > 0)] - c_1) - c_2$$
$$\Leftrightarrow c_1 > \frac{p_1}{p_1 - p}(c - pE[\pi_0(s_1 > 0)] + p^2 E[\pi_0(s_1 > 0)])$$

(B9)

By analyzing the first and second derivatives of the above boundary conditions (similar to the one-stage problem), we found that the parameter space (p, c, $E[\pi_0(s_1 > 0)]$) could be divided into three regions:

$$\begin{array}{ll} \text{Region \#1(low overall cost).} & c < (p - p^2)E[\pi_0(s_1 > 0)] \\ \text{Region\#2(moderate overall cost).} & (p - p^2)E[\pi_0(s_1 > 0)] < c < pE[\pi_0(s_1 > 0)] \\ \text{Region\#3(high overall cost).} & c > pE[\pi_0(s_1 > 0)] \end{array} \tag{B10}$$

The behaviors of the boundary conditions differ dramatically across these regions (have different signs for first and second derivatives). The value of p_1 (and its interaction with p, c) only affects the boundaries (c_1) quantitatively, i.e., changes the relative positions of the boundaries but not the shape. The results are represented in Figure 5.