1 Project Expenditure Profile

1.1 Fundamental Expenditure Behavior

The underlying distribution of expenditures (independent of cost escalation and schedule delay), $I_0(t)$, is assumed to be sinusoidal:

$$I_0(t, d) = K \frac{\pi}{2d} \sin\left(\frac{\pi t}{d}\right), \tag{1}$$

where d is the total duration of the project (in years) and K is the initial estimate of the total final project cost.

For d held constant, the cumulative expenditures on the project at time t, $E_0(t)$, is the integral of the incremental spend function:

$$\begin{aligned} E_0(t) &= \int_0^t I_0(t')\,\mathrm{d}t' \\ &= -\frac{K}{2}\left[\cos\left(\frac{\pi t}{d}\right) - 1\right] \end{aligned} \tag{2}$$

These functions have the feature that evaluating E_0 at the end of the project (time t equal to d, the project duration) reproduces the total project cost:

$$E_0(d) = K. \tag{3}$$

1.2 Characterization of Cost and Schedule Escalation Probability Distributions

Historical data on cost escalation and schedule slip was obtained from the 1985 EIA report "An Analysis of Nuclear Power Plant Construction Costs". The report provides information on anticipated project completion dates at project start (0% completion), 25%, 50%, 75%, 90%, and final project completion. The cost escalation and schedule slip that occurs between project stages is likely to be of significant interest. However, as these outcomes are expected not to be probabilistically independent (that is, the outcome in one period changes expectations about outcomes in another period), in the current version of the model total overall escalation multipliers are used.

These slip multipliers (ϵ) are calculated from the historical data by:

$$\epsilon_{cost} = \frac{\text{Overnight Cost (OC) at 90\% Completion} - \text{OC at 0\% Completion}}{\text{OC at 0\% Completion}} - 1 \tag{4}$$

$$\epsilon_{schedule} = \frac{\text{Actual Completion Date} - \text{Project Start Date}}{\text{Expected Initial Completion Date} - \text{Project Start Date}} - 1 \tag{5}$$

The R library fitdistrplus was used to fit the empirical data to exponential, gamma, and normal distributions. The dataset itself also needed to be pretreated before analysis, due to the presence of negative values and outliers.

Exponential and gamma distributions cannot produce samples of value less than or equal to zero, so negative entries in the empirical data needed to be removed or accounted for. It was possible that these negative values were themselves outliers, so in one case negative values were simply removed from the dataset. In the other case, the entire dataset was shifted upwards by a factor of $|min\{data\}| + 0.001$ prior to fitting the distribution. Subsequent samples from these shifted-range distributions would have the same shift factor subtracted out in order to recreate the original distribution with potential for negative-valued outputs.

Outliers were determined by examining the histogram for the cost and schedule ϵ datasets, and removing values that appeared to create isolated "islands" of high values, as in the following images:
[image here]

In all, ten distributions with different dataset treatments were examined; Table 1 and Table 2 show the results. Bolded lines indicate members of distribution families with the smallest parameter errors—these were considered in greater detail.

1.3 Schedule Slip

For Eq. (2) to be valid, d must be computed ahead of time and not changed during the course of the project. (On-the-fly schedule slip will be implemented in future versions.)

Each project has a statically-defined initial schedule estimate, d_0, pulled from the reactor project profile. The final duration d is generated by sampling a schedule delay factor from the previously-parameterized gamma distribution.

$$d = [1 + \Gamma(2.09296, 0.50410)]\, d_0 \tag{6}$$

1.4 Cost Escalation

1.4.1 Basic Escalation Model Form

For most historical projects, cost escalation occurred in all stages of the project and became more severe over time. The new incremental and cumulative expenditure functions are named $I_e(t)$ and $E_e(t)$, respectively. The behavior of these functions is predicated on an escalated total final project cost at time d:

$$E_e(d) = (1 + \epsilon)\, E_0(d) \tag{7}$$

The final total escalation factor, ϵ, is calculated by:

Table 1: Cost escalation distribution candidates.

Distribution	Negative Value Treatment	Outliers	Parameter Type	Value	Error
Exponential	Removed	Retained	Rate	0.58281	0.06775
Exponential	Removed	Removed	Rate	0.67580	0.08077
Exponential	**Range Shifted**	**Retained**	**Rate**	**0.51810**	**0.06023**
			Shift	**0.21431**	
Exponential	Range Shifted	Removed	Rate	0.68499	0.08129
			Shift	0.001	
Gamma	Removed	Retained	Shape	2.53879	0.39305
			Rate	1.47949	0.25321
Gamma	Removed	Removed	Shape	3.72438	0.60346
			Rate	2.51692	0.43659
Gamma	Range Shifted	Retained	Shape	3.30592	0.51841
			Rate	1.71275	0.29005
			Shift	0.21431	
Gamma	**Range Shifted**	**Removed**	**Shape**	**2.35959**	**0.37145**
			Rate	**1.61611**	**0.28339**
			Shift	**0.001**	
Normal	Retained	Retained	Mean	1.71581	0.15555
			σ	1.33809	0.10999
Normal	**Retained**	**Removed**	**Mean**	**1.45888**	**0.09143**
			σ	**0.77042**	**0.06465**

Table 2: Schedule escalation distribution candidates.

Distribution	Negative Value Treatment	Outliers	Parameter Type	Value	Error
Exponential	Removed	Retained	Rate	0.98296	0.11427
Exponential	Removed	Removed	Rate	1.03675	0.12218
Exponential	**Range Shifted**	**Retained**	**Rate**	**0.94779**	**0.10944**
			Shift	**0.05199**	
Exponential	Range Shifted	Removed	Rate	0.99738	0.11673
			Shift	0.05199	
Gamma	Removed	Retained	Shape	2.59712	0.40257
			Rate	2.55296	0.43647
Gamma	Removed	Removed	Shape	2.84002	0.44833
			Rate	2.94449	0.50838
Gamma	**Range Shifted**	**Retained**	**Shape**	**2.09296**	**0.31831**
			Rate	**1.98372**	**0.34072**
			Shift	**0.05199**	
Gamma	Range Shifted	Removed	Shape	2.21949	0.34335
			Rate	2.21368	0.38408
			Shift	0.05199	
Normal	Retained	Retained	Mean	1.00309	0.07611
			σ	0.65911	0.05382
Normal	**Retained**	**Removed**	**Mean**	**0.95064**	**0.06841**
			σ	**0.58447**	**0.04837**

$$\epsilon = \Gamma\left(\alpha, \beta\right),$$

where Γ is the gamma distribution, α is the shape parameter, and β is the rate parameter. At present, values of $\alpha = 1.2$ and $\beta = 0.6$ are being used.

This model escalates the fundamental expenditure profile $I_0(t)$ with an exponential escalator term, $e^{\alpha t}$. An appropriate exponential growth rate α must be determined, in order to reproduce the desired value of $E_e(d)$:

$$\int_0^d I_0(t)\, e^{\alpha t}\, \mathrm{d}t = E_0(d)\ (1+\epsilon) \tag{8}$$

1.4.2 Determining the Appropriate Escalation Rate Factor, α

From Eq. (8), evaluating the integral and some algebraic manipulation produce the analytically unsolvable expression:

$$e^{\alpha d} - \frac{2(1+\epsilon)\alpha^2 d^2}{\pi^2} = 2(1+\epsilon) - 1 \tag{9}$$

The bisection method is an efficient way to evaluate this expression for alpha, using probabilistically-generated values of ϵ and d. To guarantee that such a search will be able to locate the appropriate value, it must be demonstrated that the following function $f(\alpha, d, \epsilon)$ is monotonic over all possible values of α, d, and ϵ:

$$f(\alpha, d, \epsilon) = e^{\alpha d} - \frac{2(1+\epsilon)\alpha^2 d^2}{\pi^2} - 2(1+\epsilon) + 1 \tag{10}$$

If f is not monotonic in the region of interest, there will be multiple solutions for α, and the search will merely locate one of them without regard to their validity. Monotonicity is proven if the partial derivative of f with respect to α is positive everywhere in its valid domain:

$$\frac{\partial f}{\partial \alpha} = de^{\alpha d} - \frac{4(1+\epsilon)d^2}{\pi^2}\alpha \geq 0, \qquad \alpha \in [0,2],\, d \in [5,15],\, \epsilon \in [0,2] \tag{11}$$

The bisection method will evaluate the fitness of varying values of α for fixed values of d and ϵ, so it is sufficient to explore the monotonicity of the partial derivative with respect to α, rather than the gradient of f.

The value ranges for d and ϵ are set in accordance with the model's limits for acceptable escalation parameters. The value range for α was set after experimentation with the resultant alpha values for a set of test d and ϵ values that spanned their respective permitted value ranges. Actual realistic values of α are expected to be substantially less than 1; however, for the sake of robustness, α is evaluated up to 2.

$\frac{\partial f}{\partial \alpha}$ was evaluated iteratively for 1000 values each of α, d, and ϵ, evenly spanning their value ranges. See figures X and Y for plots.

[plots forthcoming]

In all tested cases the value of the partial derivative was greater than zero. Therefore, $f(\alpha, d, \epsilon)$ is monotonic and the binary-search approach for calculating appropriate α values is valid.