Friction and Decision Rules in Portfolio Decision Analysis

Published in *Decision Analysis*, articles in advance, Feb. 26, 2021.

DOI: https://doi.org/10.1287/deca.2020.0421

Gary J. Summers, PhD
President, Pipeline Physics LLC
gary.summers@PipelinePhysics.com

Abstract

In portfolio decision analysis, features comprise the objectives, alternatives, physics, and information that define a decision context. By modeling features, decision analysts forecast the expected utilities of the alternatives. A model is complete if it contains all the features. A model is well-calibrated if it correctly predicts the probability distributions of each alternative's utility, while ill-calibrated models, like those that suffer the optimizer's curse, do not. Friction identifies qualities of a situation that prevent decision analysts from creating complete, well-calibrated models. When friction is significant, can maximizing expected utility be a suboptimal decision rule? Is satisfying decision theory's axioms a necessary or sufficient condition for good decision making? Can rules that violate the axioms outperform rules that satisfy them? A simulation study of how unbiased, imprecise forecasts of payoffs affect project selection finds that, for the example tested, the answers are yes, no, and yes, which suggests that further studies of friction may be worthwhile. Discussions of friction bookend the study, starting the paper by defining friction and concluding by presenting three frameworks, each one from a different field of study, that provide mathematical tools for studying friction.

Keywords: Optimizer's curse, portfolio optimization, modeling, friction, decision rules

Friction and Decision Rules in Portfolio Decision Analysis

When selecting projects for product development, portfolio decision analysis (PDA) recommends building a model to forecast the expected utility of potential portfolios and then selecting the portfolio that maximizes the forecasted expected utility. The process should satisfy the decision theory's axioms (Savage 1954; von Neumann and Morgenstern 1953). Salo et al. (2011, p. 4) describe that, "[Decision] theory can be viewed as the foundation of PDA in that it postulates axioms that characterize rational decision making and enables the development of functional representations for modeling such decisions."

However, practitioners often apply heuristics approaches, such as scoring models, strategic buckets, and bubble charts (Cooper et al. 1998). In the pharmaceutical industry, which leads other industries in applying PDA, Kloeber (2011, p. 284) describes the limited diffusion of PDA. "Even though methods were introduced and DA [decision analysis] processes were installed in several large pharmaceutical companies as early as 1985, many other companies either failed in their attempts to internalize decision analysis concepts, such as Decision Quality, or never attempted to introduce these concepts." Likewise, describing the field of project portfolio management for product development, Kavadias and Chao (2008, p. 136) state, "Several tools and theories have been developed by different constituencies, resulting in an interesting dichotomy: a collection of rigorous analytical efforts with minimal adoption and minimal practical impact and a variety of managerial frameworks grounded in individual case studies with widespread impact but little theoretical foundation." Why do practices differ from PDA's theory and prescriptions?

Section 1 starts to sketch a hypothesis by defining friction, and in the process, it raises questions about decision making. When a model is ill-calibrated (as Section 1 defines), can

maximizing expected utility be a suboptimal decision rule? Is satisfying decision theory's axioms necessary or sufficient for good decisions? Can rules that violate the axioms outperform rules that satisfy them? For a germinal study of these questions, Section 2 introduces a simple PDA model to explore. Section 3 presents three decision rules that, within the model, compete to maximize portfolio value. Section 4 presents the competition's results and explains the answers to the above questions, which for the example are yes, no, yes. To enhance readability, the derivation of the model's parameters, which requires a detailed analysis of empirical data, occurs in an appendix. Section 5 briefly mentions mathematical models that may found theoretical analyses of friction.

Section 1: Features and friction

To investigate the baffling bifurcation of PDA theory and practice, consider PDA's cousin: modern portfolio theory (MPT). The quadratic objective function in mean-variance optimization is sensitive to imprecise and inaccurate information, called estimation errors. As Michaud (1998, p. 3) describes:

> "In practice, the most important limitations of MV [mean-variance] optimization are instability and ambiguity. MV optimizers function as a chaotic investment decision system. Small changes in input assumptions often imply large changes in the optimized portfolio. Consequently, portfolio optimality is often not well defined. The procedure overuses statistically estimated information and magnifies the impact of estimation errors. It is not simply a matter of garbage in, garbage out, but, rather, a molehill of garbage in, a mountain of garbage out. The result is that optimized portfolios are 'error' maximized and often have little, if any, reliable investment value. Indeed, an equally

weighted portfolio may often be substantially closer to true MV optimality than an optimized portfolio."

Strategies for alleviating this problem are surprising. One approach uses coarse data, filling the optimization's correlation matrix, except for the cells on the diagonal, with the same value, the overall mean, calculated by averaging all the assets' pairwise correlations. Elton and Gruber (1995, p. 169) describe this practice's success, "Tests have been performed using three different samples of stocks over a total of four different time periods. In every case, the use of the overall mean model outperformed the single-index model, the multi-index model, and the historical correlation matrix." Another strategy replaces mean-variance optimization with linear programming (Michaud 1989, 1998), eliminating the model's sensitivity but at the cost of using the wrong physics. A third strategy adds constraints to MV optimization, especially constraints to prohibit short selling (Michaud 1989), thus contradicting an implication of optimization theory that expanding the solution set can never be harmful. Finally, DeMiguel et al. (2009) tested the $1/n$ heuristic, a simple rule that allocates investment equally over assets, against 13 optimization models, with 3 models using Bayesian strategies. Competing over seven datasets, none of the optimization models consistently outperformed the $1/n$ heuristic in out-of-sample performance. How can coarse information, exploring fewer alternatives, and incorrectly modeling the physics improve performance? How can the $1/n$ heuristic outperform optimization while ignoring all information about return and risk?

Several definitions help us to propose answers. Keeney (1992) describes a *decision context* as defined by two components. One is a set of fundamental objectives that embody stakeholders' preferences, naming the values the stakeholders wish to satisfy, while providing attributes and scales for measuring satisfaction, counted in units of utility. The other is the full

set of alternatives with which decision-makers may realize the objectives. This set includes all possible alternatives, even ones that are not explicitly stated or evaluated, such as all the solutions an algorithm skillfully avoids when solving the traveling salesman problem. Connecting the alternatives to the objectives are phenomena and physics, modeled with mathematics, variables, and parameters, through which alternatives satisfy the objectives. Information describes the state of the decision environment. *Features* refer to the entire decision context: the identified objectives, the attributes and scales that measure utility, and the alternatives, physics, and information.

A model is *complete* if it includes all the features of a decision context. Over time, the PDA literature has studied more features, facilitating a march towards complete models. Heidenberger and Stummer (1999), Kavadias and Chao (2008), and Salo et al. (2011) summarize this literature. A model is *well-calibrated* if it correctly predicts the utility (deterministic situation) or probability distribution of utility (stochastic situation) of every alternative. A model that is not well-calibrated is *ill-calibrated*.

We define *friction* as issues that make a model incomplete or ill-calibrated, such as the difficulties of identifying and modeling objectives, alternatives, and physics and of interpreting information. Several studies detail these difficulties (Berkeley and Humphreys 1982; Brown 1994; Clemen 2008; French 1995; Frisch and Clemen 1994). Additionally, friction includes the biases, modeling errors, and decision errors that arise from cognitive heuristics (Einhorn and Hograth 1981; Slovic et al. 1981) and from the bounded rationality of individuals (Gigerenzer 2008; Simon 1976) and organizations (Cyert and March 1963; Jones 1999). Even for unboundedly rational decision-makers, information can cause friction, as modern portfolio theory illustrates, showing that noisy data propagating through nonlinear equations can make

models ill-calibrated. This definition may seem overly broad, containing any disturbance that might adversely affect modeling, but the definition, purposefully, is as broads as the situations that decision analysts face.

Friction is important because it can cause a decision-maker to select a suboptimal alternative, which is an alternative that has less expected utility than the optimal alternative. Mindful of these costly errors, consider decision theory's prescription to maximize expected utility, as forecasted by a model. If a model is ill-calibrated, maximizing the forecast of expected utility, although appearing like two-steps forward, could be a step or two backward. Recognizing this problem and anticipating the questions that closed this paper's introduction, Frisch and Clemen (1994, p. 48) wrote, "If people conform to utility theory, their decisions are internally consistent. It is not clear, however, that internal consistency (i.e., consistency with the axioms of utility theory) is either a necessary or a sufficient criterion for good decision making."

The term friction parallels its use in transaction cost economics (Williamson 1989), which studies how decision-makers design governance structures, such as contracts, to minimize the risks and costs created by market imperfections, called friction. Causes of friction include bounded rationality and indiscernible contingencies, which limit the effectiveness of contracts because contracts cannot cover all contingencies.

Two qualities of models can increase the impact of friction: (1) the physics that relate alternatives to objectives, such as the nonlinearities in MPT's mean-variance optimization, and (2) the features added to a model. Focusing on the second quality, some features mitigate friction, like decomposition, which improves judgments and diminishes the remaining error via weighted sums (Ravinder et al. 1988). Other features may magnify errors, such as project

interactions in PDA models, which can cause errors in one project's evaluation to affect decisions about other projects.

Following the economist's rule, we might add a feature to a model if the benefit it produces exceeds its costs. If we consider only modeling costs, such as the cost of gathering information, we will create a sophisticated, feature-rich model. This approach implicitly assumes that models are well-calibrated. In contrast, if we consider the cost of decision errors caused by friction, our models will omit alternatives, objectives, physics, and information that, when included, cause more harm than good, and these omissions will make models simpler. Generally, past some point of diminishing returns, adding features creates opportunities for modeling errors, heightens the need to simplify a model's mathematics, and requires additional information, thus exposing a model to additional bias and noise while providing pathways that propagate errors through the model to the objectives.

The above definitions present a three-step sequence that affects a decision: (1) friction causes, (2) incomplete and possibly ill-calibrated models, which can cause (3) costly decision errors. A large literature addresses the first step by helping practitioners to build models that minimize friction (Keeney 1992; Montibeller and von Winterfeldt 2015; Spetzler et al. 2016). Some research explores the transition from step 1 to step 2, studying how errors propagate through a model to produce ill-calibrated results (Clemen and Winkler 1985; Lindley 1986; Ravinder et al. 1988). Research exploring the transition from step 2 to step 3, revealing how incomplete and ill-calibrated models affect outcomes, includes studies of the value of information (Keisler 2004; Zan and Bickel 2013), the optimizer's curse (Sections 2 and 3 present a definition and citations), and robust PDA (Baker et al. 2020; Hassanzadeh et al. 2014; Kettunen and Salo 2017; Liesiö et al. 2007, 2008; Liesiö and Salo 2012; Thomas and Liesiö 2016;

Vilkkumaa et al. 2014, 2015). To this sequence, we might add a fourth step, a feedback loop from a decision's results to step 1's friction. If a decision produces unsatisfactory results, studying the results may identify the cause, either bad luck, as emphasized by decision analysis's distinction between good decisions and results, or friction that made a model imperfect and caused a decision error. By studying results, we might assign probabilities to these two causes.

Recall, the questions that concluded the introduction. Given an ill-calibrated model, can maximizing expected utility be a suboptimal decision rule? Is satisfying decision theory's axioms necessary or sufficient for good decisions? Can rules that violate the axioms outperform rules that satisfy them? We now explore these questions with a simple example, seeking to learn whether answers of yes, no, and yes are possible.

Section 2: Modeling payoffs, forecasts, and probabilities of technical success

To explore the questions and study PDA, we generate a set of projects and randomly assign each project a payoff, an unbiased but imprecise forecast of the payoff, and a well-calibrated probability of technical success. Then three selection rules compete to create portfolios that maximize value. Section 2 presents the models of payoffs, imperfect forecasts, and probabilities of technical success. Section 3 presents the selection rules. Section 4 presents the competition's results. An appendix derives the model's parameters, thus placing these long, detailed explanations last.

Payoffs and forecasts

Variables

X: The distribution of projects' payoffs, measured as the NPV of the profits a project produces, if launched, discounted to the project's launch date.

V: A normal random variable used to create a lognormal distribution of payoffs, $X = \exp(V)$.

Y: A normal random variable that creates forecasting errors, $Y \sim N(0, \sigma_Y^2)$.

Z: A probability distribution of forecasts, as specified by a forecasting model. The forecast, Z, combines each payoff from X with a random sample from Y.

Forecasting models

$Z = X + Y$: constant absolute error model (CAE).

$Z = \exp(V + Y)$: lognormal error model (LN).

$Z = X + XY$: constant relative error model (CRE).

Parameters

δ: The skew of X.

ε: The percentage of a forecast that is noise (equation 1).

$1 - \varepsilon$: The percentage of forecast that is signal.

Value of parameters used in the simulations

X: The distributions of payoffs are $X = \exp(V)$, with, $V \sim N(-0.049, 0.314^2)$, $V \sim N(-0.152, 0.551^2)$, and $V \sim N(-0.361, 0.85^2)$.

δ: $\delta \in \{1, 2, 4.18\}$, produced by the above distributions of X. Relevant references: Grabowski et al. (2002); Steedman et al. (2018).

ε: For each distribution of X, the value of σ_Y is set to produce $\varepsilon \in \{20\%, 36\%, 50\%, 61\%, 70\%, 80\%\}$. Relevant references: Cha et al. (2013).

Pharmaceutical scenario: $\delta = 2$, $\varepsilon > 66\%$.

Table 1: The model of projects' payoffs and forecasts (Section 2.1) and parameter values (Appendix) used in the simulations (Section 4).

Probabilities of technical success

Technically sound projects survive development to achieve launch.

Technically flawed projects fail in development.

SDT model: The overlapping distributions of Figure 1 create a signal detection theory model and produce well-calibrated probabilities of technical success for any level of technical uncertainty (Macmillan and Creelman 2005).

Variables

π_E: The fraction of evaluated projects that are technically sound, also called the base rate.

v: A project's score.

$\varphi_F(v) \sim N(\mu_F, \sigma_F^2)$: The probability distribution of the scores of flawed projects.

$\varphi_S(v) \sim N(\mu_S, \sigma_S^2)$: The probability distribution of the scores of sound projects.

$p = \Pr(\text{sound}|v)$: The probability that a project is sound given its score (equation 2). This probability is well-calibrated.

AUC: The area under the ROC curve (equations 3 and 4). The ROC curve plots the true-positive rate (vertical axis) against the false-positive rate (horizontal axis), or equivalently, it plots sensitivity against one minus specificity (Fawcett 2006; Gönen 2007; Macmillan and Creelman 2005).

π_D: The fraction of development projects that are technically sound.

k_D: Development cost, set to a percentile of X, $P_a = F_X^{-1}(a)$.

Parameter values

$\varphi_F(v)$: $N(0, 1^2)$.

$\varphi_S(v)$: $N(1.19, 1^2)$ and $N(1.81, 1^2)$.

AUC: The values of $\varphi_F(v)$ and $\varphi_S(v)$ are selected to produce $AUC \in \{0.800, 0.900\}$. Relevant references: Chen et al. (2015); Lindborg et al. (2014); Lo et al. (2019).

π_E: $\{35\%, 53\%\}$. Relevant references: Senn (2007); Lo et al. (2019); Arrowsmith (2011); Hay et al. (2014).

k_D: $\{P_5, P_{25}\}$. Relevant references: Grabowski (2002).

Table 2: The model of projects' probabilities of technical success (Section 2.2) and parameter values (Appendix) used in the simulations (Section 4).

Section 2.1: Modeling payoffs and forecasts

Each project that survives development achieves launch and produces a profit that is an independent draw of a random variable, X. Managers forecast this value but suffer imprecision produced by a normal random variable, $Y \sim N(0, \sigma_Y^2)$, that is independent of X. Together, X and Y produce forecasts, Z, and we seek a function that combines them to produce (1) imprecise but unbiased forecasts that (2) scale with a market's size, so that common errors, such as over- or underestimating market share or price, produce larger forecasting errors for larger markets. (Table 1 summarizes the model of payoffs and forecasts.)

Two possible models come from studies of the optimizer's curse, a phenomenon that describes how selection routinely creates overvalued, suboptimal portfolios (described in Section 3). Several studies of the curse use a *constant absolute error model* (CAE), $Z = X + Y$, which yields unbiased errors, but the "average" size of the errors, σ_Y, is the same for all values of X (Harrison and March 1984; Smith and Winkler 2006; Chen and Dyer 2009; Jorgensen 2013; Kettunen and Salo 2017; Vilkkumaa et al. 2014). Studies of the optimizer's curse in the oil and gas industry (Schuyler and Nieman 2008; Chen and Dyer 2009), and of the value of information in project selection (Keisler 2004), present *lognormal error models* (LN), $Z = \exp(V + Y)$, where $X = \exp(V)$ and Y is independent of V. The errors in this model scale with X, but they are optimistically biased. From the formula for the mean of a lognormal distribution, and recalling that the mean of Y is zero, $E[Z] = E[\exp(V)]E[\exp(Y)] = E[\exp(V)]\exp(\sigma_Y^2/2)$. Forecasts create an optimistic bias, $E[Z] > E[X]$, because $\exp(\sigma_Y^2/2) > 1$ for $\sigma_Y > 0$.

The CAE and LN models each possess one of the qualities we seek, but neither possesses both qualities together. Therefore, this study uses a *constant relative error model* (CRE), $Z = X + XY = X(1 + Y) = XY'$, where $Y' \sim N(1, \sigma_Y^2)$. The CRE model produces unbiased forecasts

because $E[Z] = E[XY'] = E[X]E[Y'] = E[X]$, and its forecasting errors grow with X because its error is XY. Relative to X, the size of the "average" error is a constant, σ_Y, since $XY/X = Y$. To model a skewed distribution of profits, which is common in many industries, let X be a lognormal distribution. Without sacrificing generality, let $E[X] = 1$, so we can characterize the lognormal distribution with only one additional parameter, its skew, δ.

The error parameter, σ_Y, impacts the CAE, CRE, and LN models differently, so we will use a metric, recommended by Keisler (2004), that maintains the same meaning in all situations. This metric is the percentage of variation in the forecast that comes from forecasting error, or in colloquial terms, the noise in the forecast. For a CRE model with lognormal X scaled so $E[X] = 1$, this percentage is:

$$\varepsilon = \frac{Var(XY)}{Var(X + XY)} = \frac{[Var(X) + 1]\sigma_Y^2}{[Var(X) + 1]\sigma_Y^2 + Var(X)} = \frac{\sigma_Y^2 \exp(\sigma_X^2)}{(\sigma_Y^2 + 1)\exp(\sigma_X^2) - 1} \qquad (1)$$

The complement, $1 - \varepsilon$, is the percentage of the forecasts that is signal.

Section 2.2: Modeling projects' probabilities of technical success

Having introduced the model of payoffs and forecasts, we address projects' probabilities of technical success. (Table 2 summarizes this model.) In the pharmaceutical industry, statisticians design clinical trials to produce desired false-positive and false-negative rates. Combining these rates with a prior probability distribution, via Bayes' law, estimates a trial's technical probability of success (Beckman et al. 2011; Chaung-Stein et al. 2011; Chen and Beckman 2007, 2009a, 2009b; Lendrem and Lendrem 2013; Lindborg et al. 2014; Patel and Ankolekar 2007, 2015; Summers 2010). By generalizing this model, we can create well-calibrated probabilities of technical success for projects. Assume projects come in two types. *Technically sound* projects will succeed in development, but *technically flawed* ones will fail. If

decision-makers evaluate technical risk with additive models, such as scoring models (Cooper et al. 1998) or simple linear models (Dawes 1979), the central limit theorem implies normally distributed scores. Figure 1 illustrates this result, which is a signal detection theory model (Macmillan and Creelman 2005). Technically flawed projects receive scores, v, from the distribution on the left, $\varphi_F(v) \sim N(\mu_F, \sigma_F^2)$, whereas technically sound projects receive scores from the distribution on the right, $\varphi_S(v) \sim N(\mu_S, \sigma_S^2)$. Let π_E, sometimes called a base rate, be the percentage of the evaluated projects that are technically sound. Bayes' law yields the probability of a project being sound as a function of its score:

$$p = \Pr(\text{sound}|v) = \frac{\pi_E \varphi_S(v)}{\pi_E \varphi_S(v) + (1 - \pi_E)\varphi_F(v)} \tag{2}$$

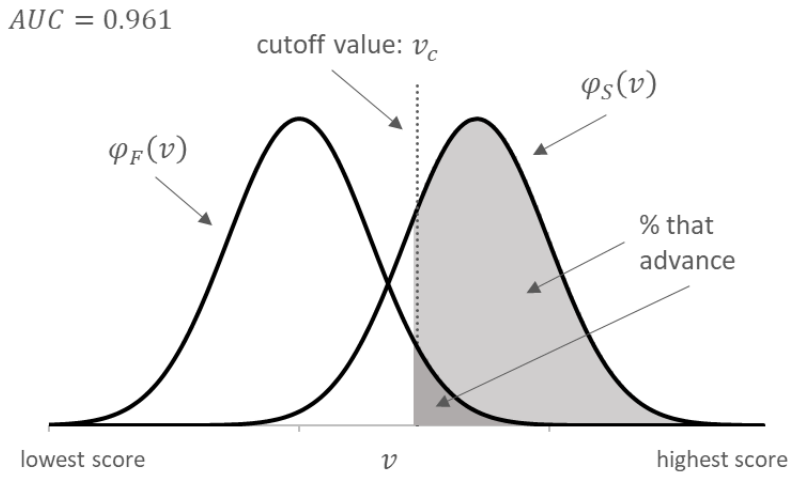


Figure 1: A signal detection theory model of scores that evaluate technical risk.

The overlap of the curves represents uncertainty, and one metric that measures it is called the area under the ROC curve, for which AUC is the standard notation (Fawcett 2006; Gönen 2007). The AUC is the probability of assigning a higher score to a randomly selected sound

project than to a randomly selected flawed project. Technical uncertainty is absent when the two distributions are entirely separate (no overlap), and then $AUC = 100\%$. Pervasive uncertainty, producing the worst evaluations, with $AUC = 50\%$, occurs when the two distributions overlap entirely, such as when they have equal means and standard deviations. Lower values of AUC identify greater technical uncertainty, whereas higher values indicate greater technical resolution. The AUC, determined by the overlap of the curves, is:

$$AUC = \Phi\left(\frac{u}{\sqrt{1+r^2}}\right) \tag{3}$$

where $u = (\mu_S - \mu_F)/\sigma_S$ and $r = \sigma_F/\sigma_S$. If a decision-maker sets a cutoff value, v_c, and advances projects with $v > v_c$, the resulting false-positive and false-negative rates give the AUC as a function of the error rates:

$$AUC = \Phi\left(\frac{z(1-\beta) - z(\alpha/2)}{\sqrt{2}}\right) \tag{4}$$

where $\beta = \Phi_S(v_c)$ is the false-negative rate, $\alpha/2 = 1 - \Phi_F(v_c)$ is the false-positive rate, and the function $z(.)$ gives the z-score of the error rates.

The final parameter in the model of development is cost. For simplicity, all projects advanced to development incur the same development cost, k_D, and to ensure consistency over the scenarios (see Appendix), the development cost is specified as a percentile of X, $k_D = P_a = F_X^{-1}(a)$.

Section 3: Three project selection methods

The simulation creates numerous projects and assigns a random payoff, forecast, and probability of technical success to each one. Three selection rules then compete by constructing portfolios to maximize value. This section introduces the project selection rules (summarized in Table 3), but it begins by considering how imperfect forecasts harm selection. Forecasting errors can cause three problems (1) overestimating portfolio value (the optimizer's curse), (2) selecting suboptimal portfolios, and (3) reducing portfolio value (Vilkkumaa et al. 2014). The optimizer's curse exemplifies these problems. Analogous to the winners curse in economics (Thaler 1988), project selection methods, seeking to maximize value, select some overvalued projects and reject some undervalued ones. On average, selection creates overvalued portfolios. Formally, for n projects with payoffs $x_1, \dots, x_n$, let $z_1, \dots, z_n$ be forecasts that are unbiased, meaning $E[z_i|x_1, \dots, x_n] = x_i$. Let i^* denote the alternative with the maximal estimated value $z_{i^*} = \max\{z_i, \dots, z_n\}$. Smith and Winkler (2006) prove $E[z_{i^*} - x_{i^*}] \geq 0$, and if there is a positive probability of selecting a suboptimal alternative, $E[z_{i^*} - x_{i^*}] > 0$. Vilkkumaa et al. (2014) extend this result to portfolios comprising a proper subset of the n projects, and for the special case of selecting projects via a cutoff value, z_c, for both the CAE model with normal X and the LN model, Chen and Dyer (2009) prove $\sigma_Y > 0$ implies $E[Z|Z > z_c] > E[X|Z > z_c]$.

Selection via expected values

EV:	For each project, forecast its expected value, pz. Set a cutoff value. Select projects with forecasts exceeding the cutoff value.
t:	The percentage of evaluated projects that exceed the cutoff value and thus advance to development (throughput).
$g(t)$:	The value created by EV as a function of throughput.
t^*_{EV}:	The optimal throughput for EV.

Selection via Bayesian adjusted expected value

BE:	Assume decision-makers know the distribution of projects' payoffs, X, the relative size of forecasting errors, σ_Y, and the error model (CRE model in this experiment). Via Bayes' law, calculate the expected payoff given the forecast, $E[X\|Z = z]$, and forecast expected value as $E[X\|Z = z]p$. Set a cutoff value. Select projects with forecasts exceeding the cutoff value.
$b(t)$:	The value created by BE as a function of throughput.
t^*_{BE}:	The optimal throughput for BE.

Selection via the two-screen method

TS:	Set cutoff values for payoffs, z_c, and for project scores, v_c. Advance projects with forecasts and probabilities that exceed their respective cutoff values.
t_M, t_T:	The percentage of projects that survive the market screen and the percentage that survive the technical screen.
$h(t_M, t_T)$:	The value created by TS as a function of the market and technical throughput.
$t_M t_T$:	The throughput produced by TS.
$t^*_M t^*_T$:	The optimal throughput for TS.

Table 3: The selection rules (Section 3) used in the simulations (Section 4).

Section 3.1: A calculation of expected value

For simplicity, assume decision-makers seek to maximize expected value, rather than expected utility. With estimates of p and z, a decision-maker might calculate a projects' expected value as pz. In the present model, where all projects have the same development cost, the decision-maker can maximize value by setting a cutoff value and advancing only those projects with expected values that exceed the cutoff. Identify this method as EV. To measure its performance, let t be the percentage of projects that advance (throughput), and let $g(t)$ be a function that gives the ex ante value of a project being evaluated for development. For a sufficiently high cutoff value, no projects advance, so $g(t = 0) = 0$. For a sufficiently low cutoff value, all projects advance, so $g(t = 100\%) = \pi_E E[X] - k_D$. Some level of throughput, t^*_{EV}, maximizes projects' ex ante value at $g(t^*_{EV})$.

The pz metric is compensatory because it balances forecasted profit with probabilities of technical success, such as advancing projects that have low values of p but high values of z. It is rational because the metric produces preference orderings of projects and portfolios that satisfy decision theory's axioms. However, despite being produced by well-calibrated probabilities and unbiased forecasts, EV produces an ill-calibrated PDA model. The problem is the optimizer's curse.

Consider a drug that is forecasted to be a blockbuster. Likely, the high value of z arises in one of two ways: (1) the compound is indeed a blockbuster and the forecasting error is small or (2) the compound is not a blockbuster and the forecasting error is large. Because of the skewed distribution of profits, the second case is much more likely, so forecasts of blockbusters are, on average, optimistic. Derived for the mathematically tractable LN model (Chen and Dyer

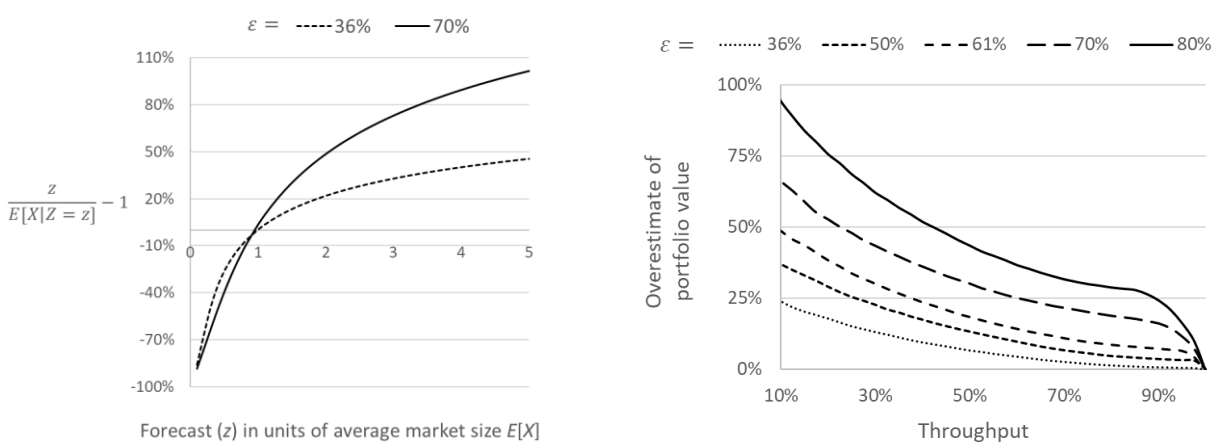
2009), the formula for $E[X|Z = z]$ shows these biases, revealing how unbiased, imprecise forecasting errors produce a striking pattern of biases when conditioned on Z:

$$E[X|Z = z] = \exp[(1 - \varepsilon)\ln(z) + \varepsilon\mu_V + \varepsilon\sigma_V^2/2] \quad (5)$$

For $\varepsilon > 0$, setting $E[X|Z = z] = z$ reveals that the expectation equals the estimate only when the estimate equals the population mean, $z = E[X]$. On average, forecasts of $z > E[X]$ are optimistically biased. For all three models (LN, CAE, and CRE) there is a value, z_m, such that if $z > z_m$, then $z > E[X|Z = z]$.

Figure 2A illustrates the biases by plotting $(z/E[X|Z = z]) - 1$ for the CRE model with $\delta = 2$ for $\varepsilon = 70\%$ (representing pharmaceuticals) and $\varepsilon = 36\%$. In pharmaceuticals, the top decile of products, which is a good definition of a blockbuster, has profits of five times the industry average (Grabowski et al. 2002). These forecasts are highly biased: $E[X|z = 5E[X]] \approx 50\%z$.

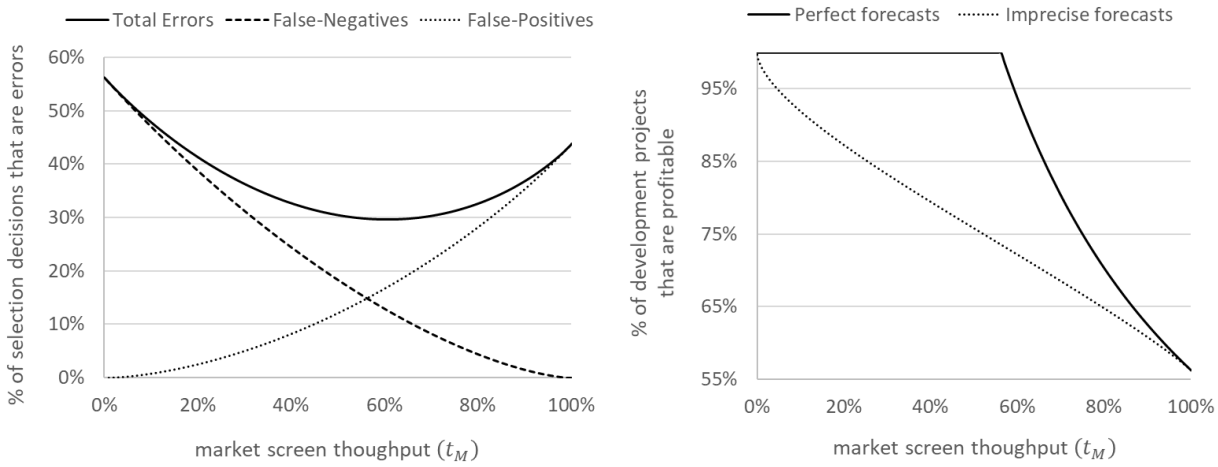
For the biases that Figure 2A displays, any nonrandom selection method produces optimistic estimates of portfolio value. Figure 2B illustrates this optimism in simulations of selection via EV for a scenario that, by some measures, matches pharmaceuticals (see Appendix). Industry-wide, 32.4% of phase 2 compounds advance to phase 3 (Hay et al. 2014), so, on average, the scenario suggests that project selection based on pz overvalues phase 3 portfolios by about 40%. Notice that throughput of 100% eliminates the bias, as optimistic and pessimistic forecasts cancel each other. When there is no selection, there is no selection bias.



Figures 2A (left) and **2B** (right): Both charts come from a CRE model where X has a skew of $\delta = 2$. Figure 2A shows the average bias in a project's forecast as a function of the forecast, z. Figure 2B shows the average bias in forecasts of portfolio value when selecting projects via EV, presented as a function of throughput.

Section 3.2: A two-screen selection method

If the forecasting imprecision causes sufficiently large estimation errors, selection with EV makes decision errors and creates suboptimal portfolios. To illustrate these errors, focus on profit forecasts. A successful development project must pay for its development plus a portion of the development cost from projects that fail, so we define a profitable project as having a payoff $x > k_D/\pi_D$, a false-negative as canceling a profitable project, and a false-positive as advancing an unprofitable project ($x \leq k_D/\pi_D$). For a high development cost, large skew and large forecasting error, Figure 3A shows that project selection errors comprise a minimum of 30% of the project selection decisions.



Figures 3A (left) and **Figure 3B** (right): Using the LN model, which produces nicely shaped curves, these figures show the selection errors produced when picking projects by their forecasted payoffs, using a cutoff value, z_c. The percentage of projects that are selected (throughput) is $t_M = 1 - F_Z(z_c)$. The parameters for this LN model were selected to match situations found in the pharmaceutical industry (see Appendix), except for the example's large skew: $\delta = 4.18$, $\varepsilon = 66\%$ and $k_D/\pi_D = P_{43.6}$.

These costly errors arise because EV trades well-calibrated probabilities of technical success for ethereal payoffs. Rather than corrupting the good technical data with the noisy market data, a selection method that uses p and z separately may produce fewer errors and thus create more value than selection via pz. Focusing on extremes: perfectly forecasted payoffs make expected values the best selection metric, but forecasts with infinite imprecision make expected values useless so that selecting based only on the probability of technical success becomes superior. A transition must occur. If it occurs for realistic levels of forecasting error, the two-screen heuristic is a valuable decision rule.

Decision-makers apply a technical screen by assigning a score or a probability of technical success to each project. (Either metric works because equation 2 establishes a one-to-one relationship between scores and probabilities with both metrics producing the same ordering of projects.) Our analysis proceeds by using the projects' scores. Decision-makers set a cutoff value, s_c, and advance projects with scores $s > s_c$. Figure 2 illustrates the fraction of sound projects and of flawed projects that advance. Before the evaluation (ex ante), the probability of a project surviving the screen, the technical throughput, is:

$$t_T = \pi_E[1 - \Phi_S(v_c)] + (1 - \pi_E)[1 - \Phi_F(v_c)] \quad (6)$$

The fraction of advancing projects that are technically sound, which is the success rate in development is:

$$\pi_D = \frac{\pi_E[1 - \Phi_S(v_c)]}{t_T} \quad (7)$$

The decision-maker performs a second selection, a market screen, by setting a cutoff value, z_c, and advancing projects with $z > z_c$. The ex ante probability of a project surviving this screen, the market throughput, is $t_M = 1 - F_Z(z_c)$.

Assuming the technical and market characteristics and evaluations are independent of each other, a project's ex ante value when selected via the two-screen method is $h(t_M, t_T) = t_M t_T(\pi_D E[X|Z > z_c] - k_D)$. The throughput that advances proposals to development is $t_M t_T$. The optimal cutoff values produce a value of $h(t_M^*, t_T^*)$. This heuristic is noncompensatory and violates decision theory's completeness axiom because, being unable to compare differences in forecasted payoffs to differences in the probabilities of technical success, it cannot produce a complete preference ordering. For conciseness, TS identifies this method.

Section 3.3: Bayesian estimates

The TS method manages uncertainty, and the EV method ignores it. The third selection method in this experiment resolves uncertainty. The forecast z provides information about an object (a case) that is a sample of a class, X. Combining case data with class data interweaves information from two perspectives to resolve uncertainty (Åstebro and Koehler 2007; Kahneman and Lavollo 1993; Kahneman and Tversky 1982; Lovallo and Kahneman 2003), and calculating $E[X|Z = z]$, via Bayes' law, accomplishes this task. The result reduces forecasting error, increases portfolio value, and, if done with knowledge of the error model, including the distributions of X and Y, eliminates the optimizer's curse as well (Vilkkumaa et al. 2014). To test this approach, called BE, assume managers calculate $E[X|Z = z]$ for each project, estimate each project's expected value as $E[X|Z = z]p$, set a cutoff value, and select projects with expected values above the cutoff. Let $b(t)$ give the ex ante expected value of a project being evaluated for development when managers use BE. The optimum throughput, t^*_{BE}, maximizes value at $b(t^*_{BE})$. (Appendix A6 derives $E[X|Z = z]$ for the CRE model.)

Section 4: Simulation experiments

To test BE, EV, and TS, the simulations use realistic high and low values of δ, ε, π_E, AUC, and k_D, gleaned from empirical data for a critical decision in drug development: selecting phase 2 compounds to advance to phase 3 clinical trials (see Appendix). The parameter values are $\delta \in \{1, 2, 4.18\}$, $AUC \in \{80\%, 90\%\}$, $\pi_E \in \{35\%, 53\%\}$ and $k_D \in \{P_5, P_{25}\}$. The parameter values for forecasting error span a full range, from exceptional to terrible forecasts, $\varepsilon \in \{20\%, 36\%, 50\%, 61\%, 70\%, 80\%\}$. Importantly, derived from Cha et al. (2013), a reasonable

forecasting error when evaluating compounds for phase 3 trials is $\varepsilon > 66\%$, suggesting a surprising low signal of $1 - \varepsilon < 34\%$ (see Appendix).

The cross product of the parameters produces 144 scenarios. For each scenario, the experiment performed at least three simulations, each one composed of 5,000 projects randomly created with the CRE model. Specifically, each project received a profit from a random draw of a lognormal X, a forecasting error from a random draw of Y, a type, flawed or sound, selected by a Bernoulli variable with probability π_E, and a score, produced from a random draw of $v = \varphi_F^{-1}$ or $v = \varphi_S^{-1}$, depending on whether a project was flawed or sound.

After the simulation creates the projects, each selection rule (EV, BE, and TS) selects projects. Of the selected projects, the technically flawed ones fail in development, but the technically sound ones succeed, launch, and produce their payoffs. Then the simulation calculates the value in each portfolio, the ex ante value of a project being evaluated, and the ROI of development. The simulation calculates these results for all cutoff values (for upcoming Figures 4A and 4B) and for each rule's optimal cutoff value (for Figure 5). The relationships presented below are robust, checked by repeating all the simulations for the LN model and the CAE and CRE models with exponential distributions of X.

Section 4.1: Exploring the three questions

Recall the questions that concluded Section 1. If a model is ill-calibrated: Can maximizing expected utility be suboptimal? Is satisfying decision theory's axioms necessary or sufficient for good decisions? Can rules that violate the axioms outperform rules that satisfy them? The competition between EV and TS offers a modicum of insight. Figures 4A and 4B present the results from a simulation with $\delta = 1$, $\varepsilon = 61\%$, $\pi_E = 53\%$, $AUC = 0.9$, $k_D = P_5$, so

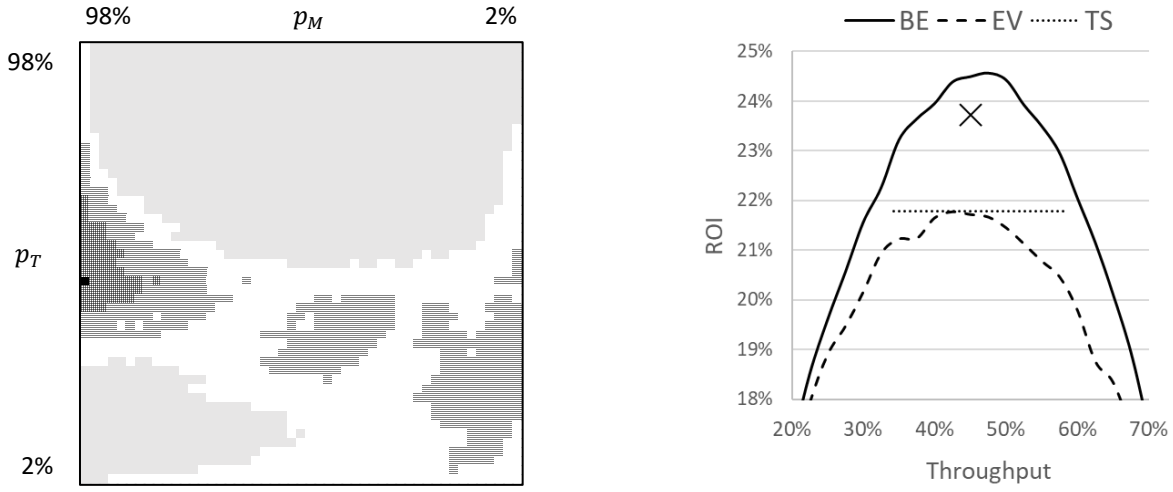
the development costs, skew of the payoffs, and forecasting errors are smaller than in the pharmaceutical industry, as they are in most industries. In Figure 4A, the horizontal axis is $98\% \geq t_M \geq 2\%$ and the vertical axis is $98\% \geq t_T \geq 2\%$ with $(98\%, 98\%)$ in the top left corner and $(2\%, 2\%)$ in the bottom right corner. Each point in the chart represents a possible selection via TS with a throughput of $t_M t_T$. The chart compares the performance of TS and EV via the difference $h(t_M, t_T) - g(t_M t_T)$. To make a measurement that adjusts to each simulation, let $N = 5\% g(t_{EV}^*)$. The solid gray area shows where EV significantly outperforms TS: $h(t_M, t_T) - g(t_M t_T) \leq -N$. The white area indicates that EV is only slightly superior: $0 > h(t_M, t_T) - g(t_M t_T) > -N$. The horizontal stripes show where TS is slightly superior: $N > h(t_M, t_T) - g(t_M t_T) > 0$. The vertical and horizontal stripes together show where TS is significantly superior: $h(t_M, t_T) - g(t_M t_T) > N$.

Gray areas fill most of Figure 4A, with an exception, which exists for nearly all the scenarios tested, of a white swoosh that extends from the upper-left area to the bottom-right corner. For low levels of ε, the swoosh is solid white, signaling that EV outperforms TS throughout this area, and thus everywhere. As ε increases, TS becomes slightly superior to EV on the extreme left side of the swoosh (horizontal lines, Figure 4A). Further increases of ε cause (1) the area where TS beats EV to grow rightward across the swoosh and (2) produce an area, at the extreme left of the swoosh, where TS markedly beats EV (horizontal and vertical lines together). Further increases in ε, to high values, create a continuous zone, from extreme left to the extreme right of the swoosh, throughout which TS beats EV.

Focusing on the white swoosh, we can identify the situations in which TS outperforms EV. The optimal cutoffs for TS, (t_M^*, t_T^*), are always within the white swoosh. When k_D is small, projects are highly profitable, so the optimal throughput is high. In these cases, (t_M^*, t_T^*) is

on the extreme left, with $t_M^* = 100\%$ (black dot, Figure 4A). As k_D increases, (t_M^*, t_T^*) travels across the white swoosh towards the bottom-right corner. Whenever (t_M^*, t_T^*) is within a stripped area, $h(t_M^*, t_T^*) > g(t_M^* t_T^*)$, and except for the borderline cases, where the methods have nearly the same performance, $h(t_M^*, t_T^*) > g(t_{EV}^*)$ as well. For example, in Figure 7A, $h(t_M, t_T) > g(t_{EV}^*)$ for $t_M \geq 90\%$ and $38\% \leq t_T \leq 58\%$.

Figures 4B compares the three selection methods by presenting the ROIs produced by BE and EV as a function of throughput, with ROIs calculated like the one for EV: $[g(t) - k_D]/k_D$. An X marks the maximum ROI produced by TS, revealing that TS's maximum ROI is closer to $b(t_{BE}^*)$ than to $g(t_{EV}^*)$. The horizontal dotted line shows the range of throughput for which $h(t_M, t_T) > g(t_{EV}^*)$, for some combination of (t_M, t_T), which in this simulation was 34% to 58%. In fact, managers can outperform EV by ignoring profit forecasts (setting $t_M \geq 100\%$) and choosing a cutoff value for the probability of technical success within the interval $43\% \leq p \leq 75\%$. This is a broad range, but if capacity constraints require less throughput, TS outperforms EV when capacity constraints place (t_M, t_T) within the area were TS is superior (Figure 4A, horizontal lines).



Figures 4A (left) and **4B** (right): For the scenario with $\delta = 1$, $\pi_E = 53\%$, $AUC = 0.9$, $k_D = P_5$, and $\varepsilon = 61\%$. Figure 4A compares TS with EV. The gray and white areas show where $h(t_M, t_T) < g(t_M t_T)$, while the stripped areas show where $h(t_M, t_T) > g(t_M t_T)$. Figure 4B shows the ROIs produced by EV and BE as a function of throughput. The X is the maximum ROI produced by TS, and the dotted horizontal line shows throughput levels where TS outperforms EV for some combination of (t_M, t_T).

Section 4.2: Which selection rule is best?

For every scenario producing an ROI between 10% and 100%, the realistic scenarios, Figure 5 presents the ROIs produced by $h(t_M^*, t_T^*)$, $g(t_{EV}^*)$ and $b(t_{BE}^*)$ as a function of ε. Each data point presents the average of at least three simulation runs, with more simulations added selectively to maintain a somewhat constant value for the maximum ROI produced when selecting projects by their true expected values, ρx. (Without this smoothing, the curves are

wavy because the skewed distributions produce substantial variation in ROI, arising from a few highly profitable projects in each simulation. Creating enough simulations to smooth this volatility requires a prohibitive amount of time.)

Highlighting the most obvious result, if made with the correct X, Y, and forecasting model, to which practitioners are not privy, BE is always best. The result occurs because BE reduces uncertainty. For example, in the CRE model, when the error in the forecasts, Z, is $\varepsilon = 80\%$, the error in $E[X|Z = z]$, which is the estimate BE uses for selection, is $\varepsilon \approx 36\%$. Figure 5 plots the ROI of BE as a function of the error in z, not the error in $E[X|Z = z]$, so we can view the ROI of BE as the maximum value one can obtain given the error in Z.

Comparing TS and EV, recall that they use the same information, that EV satisfies decision theory's axioms and explicitly maximizes forecasts of expected value, while TS does not forecast portfolio value and violates decision theory's completeness axioms. For low-cost scenarios and small skews (Figure 5, rows 1 and 2), TS outperforms EV at surprisingly low values of market uncertainty, in one case when $\varepsilon \geq 36\%$. Now compare TS's performance for P_5 with P_{25}. When $k_D = P_{25}$ the optimal market throughput is $t_M^* < 100\%$ for most values of ε. In these situations, TS is using market information, so its ROI decreases as ε increases. However, as ε increases, the optimal cutoff value for the market screen, z_c^*, decreases, which means the optimal market throughput, t_M^*, increases, making TS depend less on market information. Eventually, as ε increases, $t_M^* = 100\%$ and TS becomes impervious to market uncertainty because it no longer uses forecasts. As ε increases further, TS maintains a constant level of performance, especially compared to BE. In contrast, EV always uses market information, so it always suffers when ε increases.

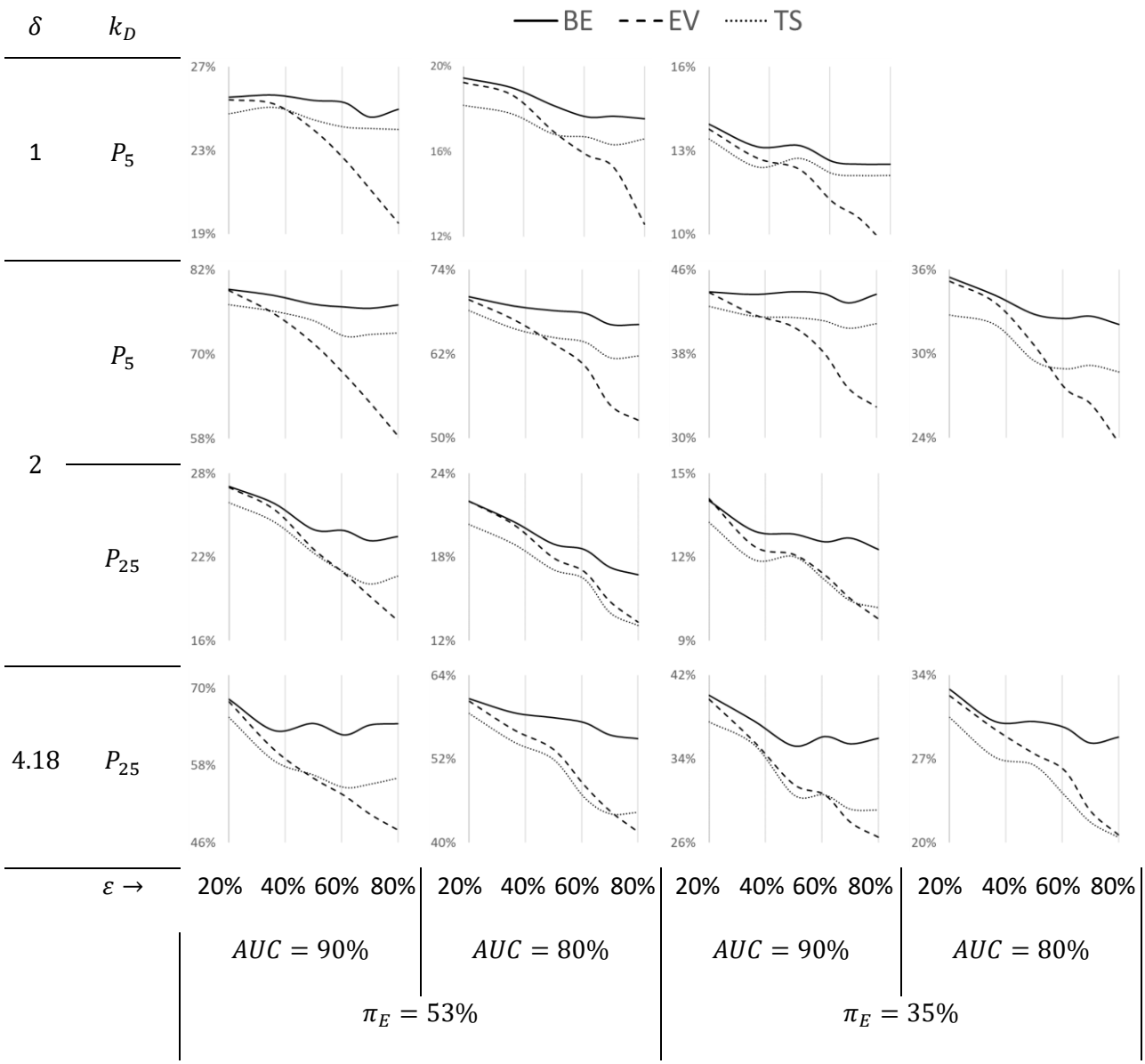


Figure 5: The ROIs produced by $b(t^*_{BE})$, $g(t^*_{EV})$, and $h(p^*_M, p^*_T)$ as a function of market uncertainty. The simulations tested values of $\varepsilon \in \{20\%, 36\%, 50\%, 61\%, 70\%, 80\%\}$, and Excel interpolated the values between them.

Both TS and BE succeed by excluding data that, while potentially useful, is sometimes harmful. For BE, as σ_Y increases, Bayes' law relies less on z and more on the distribution of X, eventually omitting z as $\sigma_Y \to \infty$ (equation 5). In contrast, EV always uses market information, so it always suffers when ε increases. The ability to exclude potentially harmful information is a quality that TS and BE have in common but that EV lacks.

Section 4.3: Do pharmaceutical companies use BE?

Do companies use BE? Citing texts on Bayesian statistics (Carlin and Louis 2000; Gelman et al. 2013), some literature describes integrating historical information into forecasts via hierarchical Bayesian models (Lenk and Rao 1990; Neelagegham and Chintagunta 1999; Pammer et al. 2000; Sultan et al. 1990). This approach uses Bayes' law when calculating z by folding class data into a model's parameters. In contrast, BE uses Bayes' law after estimating z. Even if the forecasting process is unbiased, each estimate is either an overestimate or an underestimate, and the forecast provides information about the probability and magnitude of each possibility (Figure 2).

Stonebraker and Keisler (2011) studied the database of a large pharmaceutical company to see how the company forecasted the net present value, if launched, of 223 drugs in development. The firm estimated NPV for forecasts for various levels of profit, such as an upside outcome, most likely outcome, and downside outcome, and estimated probabilities for each scenario, thus forecasting z and estimating $f(Z|X = x)$. No subsequent Bayesian adjustments were described.

Meanwhile, based on Cha et al. (2013), sell side analysts' forecasting errors of peak sales is $\varepsilon > 66\%$ (see Appendix). Errors this large suggest the analysts' estimates are predominantly z rather than $E[X|Z = z]$. In simulations with high uncertainty, $\varepsilon = 80\%$, and X having a reasonable skew for pharmaceuticals, BE reduced the standard deviation of forecasting errors by 67% to produce $\varepsilon \approx 36\%$. If the analysts calculated BE, the result was uninspiring.

Why might practitioners forgo BE? The reasons parallel the sources of friction that Section 1 described: information and bounded rationality at both the individual and organizational levels. Focusing on information, in practice X, Y and the error models must be ascertained, and estimating X is particularly difficult if the payoffs are changing over time. Adding these features may introduce bias and additional noise into the model, perhaps worsening the model's calibration. Bounded rationality from organizational needs and processes may preclude BE as well. Stonebraker and Keisler (2011) report that marketing departments estimate z for numerous scenarios, and portfolio management groups combine the scenarios with probabilities. Having "outsourced" the forecast, the portfolio management team may not know that additional adjustments might improve the forecasts. Finally, individual bounded rationality may hinder the estimation of X. Åstebro and Koehler (2007) studied expert evaluations of new product proposals and found that experts committed the cognitive error of ignoring base rates when estimating probabilities of success (Tversky and Kahneman 1982). Base rates are class information, as is X.

Considering friction, unless the Bayesian calculation can sufficiently reduce ε, TS is the best decision rule. Possibly, a simpler estimate of BE, made by taking a weighted average of z and an estimate of $E[X]$, could reduce forecasting errors without introducing too much noisy

information or mistaken assumptions. If successful, this heuristic adjustment could make the explicit maximization of expected value into the best decision rule.

Section 5: Three mathematical frameworks for theorizing about friction

Section 4 presents an example in which realistic amounts of imprecise information turn the explicit maximization of expected value into a suboptimal decision rule. Friction matters. While intriguing, more compelling study requires theoretical analysis, perhaps deriving the normative implications of friction. The following three articles provide formal frameworks that may support such efforts.

Heiner (1983, 1986, 1988) incorporates bounded rationality into decision-theoretic models by using the conditional probabilities that comprise value of information calculations, albeit with two alterations of conventional practice. First, rather than estimate the value of improved information, Heiner calculates the cost of decision errors. Second, he applies the calculations to decision errors arising from bounded rationality. Heiner shows that boundedly rational decision-makers must mitigate the cost of decision errors by using fewer alternatives and less information, which simplifies decision frames. The resulting optimal behavior is coarse, meaning it is less variable than the behavior of an unboundedly rational decision-maker.

Al-Najjar and Pai (2014) address framing in their study of decision errors caused by small datasets. They view decision-makers as frequentists who estimate a model's parameters by using data. For situations like curve fitting and categorization, they show that the total cost of decision errors is the sum of two costs: the costs caused by having an imperfect frame plus the cost of overfitting. Adopting a coarse (simpler) frame increases the cost arising from the imperfect frame but reduces the cost caused by overfitting. As data becomes scant, which

exacerbates overfitting, the trade-off becomes beneficial. The behaviors arising from coarse frames are less variable than the behavior recommended by optimizing with a fine frame.

Inspired by Al-Najjar and Pai (2014), consider classifying decision errors into *errors of omission* (the features missing from an incomplete model) and *errors of commission* (such as modeling errors and sensitivity to imperfect information and imperfect models). Decision analysts may strive to build models that minimize the total cost of decision errors, measured as expected utility, calculated as:

$$\begin{array}{c}\text{Expected cost of}\\ \text{decision errors}\end{array} = \begin{array}{c}\text{Expected cost from}\\ \text{errors of omission}\end{array} + \begin{array}{c}\text{Expected cost from}\\ \text{errors of commission}\end{array}$$

This additive relationship of errors exists with decision trees. Let the forecast of each terminal branch's payoff be its true value plus an error term, $a_i = a_i^T + a_i^\epsilon$, and the estimated probability of each terminal branch occurring be its true probability plus an error term, $w_i = w_i^T + w_i^\epsilon$, producing a forecasted expected utility of $u_1 = \sum w_i a_i = \sum (w_i^T + w_i^\epsilon)(a_i^T + a_i^\epsilon)$, as described in Ravinder et al. (1988). Rearranging these terms produces $u_1 = \sum w_i^T a_i^T + \sum w_i^\epsilon a_i^T + \sum w_i^T a_i^\epsilon + \sum w_i^\epsilon a_i^\epsilon$, where the first summation is the true expected utility and the remaining summations are estimation errors caused by friction. To improve decision making, we might add features to the model, thereby expanding the decision tree, creating new terminal branches, and producing a new estimate of utility, u_2. In the difference $u_2 - u_1$, the true expected utility, $\sum w_i^T a_i^T$, appears in both terms and drops out of the subtraction. The remaining terms show how the modeling change affected the errors. Improvements from reducing the errors of omissions shrink some error terms, while errors of commission, which occur if the added features incorporate additional friction into the model, increase other error terms. The optimal model trades errors of omission for errors of commission, or vice versa, to minimizes the cost, measured in expected utility, of decision errors.

In a third framework, Bookstaber and Langsam (1985) describe a single decision from three perspectives, with each perspective producing an optimization model. First, a decision-maker knows only a proper subset of the variables that define the environment. Ignorant of some variables, the decision-maker creates an incomplete model, thereby hamstringing the decision-maker with unmodeled uncertainty. Second, an omniscient observer knows the entire state space and thus models all uncertainty with probabilities. Third, the optimal decision rule, created by the omniscient observer, tailored to the decision-maker's ignorance but without providing the decision-maker with new information, produces optimal decisions while laboring under the decision-maker's ignorance. Comparing the three optimizations yields insights, including the following. The optimal behavior for the decision-maker (optimization 3) is less variable than the behavior produced by optimizing while assuming the decision maker's incomplete model is complete (optimization 1).

The introduction of the omniscient observer's perspective is provocative. By optimizing as if the incomplete model were complete, the decision-maker believes that the resulting actions are optimal. However, from the omniscient observer's perspective, those decisions may be suboptimal because, even if the decision-maker adheres to decision theory's axioms, the decision-maker may have an erroneous preference ordering. Meanwhile, the decision rule in the third optimization, the decision-maker's optimal rule, may look suboptimal to the decision-maker, possibly even violating decision theory's axioms. To the omniscient observer, the rule is optimal and consistent with decision theory's axioms.

Conclusion

Friction can make models ill-calibrated so that the models incorrectly forecast expected utility. If practices like debiasing and decision quality sufficiently mitigate friction, decision making should follow decision theory's traditional prescriptions: adhere to its axioms and select the alternative that maximizes a model's forecast of expected utility. However, what happens if friction is unavoidably impactful, creating significant errors in a model's estimates? Can maximizing expected utility be a suboptimal decision rule? Is satisfying decision theory's axioms necessary or sufficient for good decisions? Can rules that violate the axioms outperform rules that satisfy them? This paper's example suggests answers of yes, no, and yes are possible and thus worthy of further study. Such studies might explain why PDA practices differ from theory, or, by developing new methods, methods that manage friction well, cause theory and practice to converge, making them look more like twins, or at least like siblings, instead of resembling distant cousins.

Appendix: Estimating the model's parameters with empirical data

A1: Estimating the skew of the payoffs, δ

To test BE, EV and TS, we need realistic ranges of δ, ε, π_E, AUC, and k_D, spanning from low to high values, gleaned where possible from empirical data. Data from the pharmaceutical industry provides some guidance, especially for the critical selection of compounds for phase 3 trials. The set of choices is the compounds that have completed phase 2 trials, which are small trials that provide initial evidence of efficacy and safety. Compounds that advance to phase 3 receive large clinical trials that produce the evidence of safety and efficacy the FDA requires for approval.

Grabowski et al. (2002) report that the top three deciles of pharmaceuticals launched from 1990-1994 produced approximately 50%, 20%, and 10% of the industry's profits, implying a lognormal distribution with $\delta = 15.6$, an absurdly large skew. Using a different approach, we could estimate δ via a key relationship: for any development cost, ROI increases with the skew of X. Steedman et al. (2018) estimate the aggregate ROI of late-stage development (phase 3) for twelve large-cap pharmaceutical firms, from 2010-2018, fell from 10.1% to 1.9%, suffering from increased costs and decreasing revenue per drug. DiMasi et al. (2003) provide cost data that compliments the revenue data from Grabowski et al. (2002), and several recent studies exist as well. Using this data, the simulations of Section 4 can find the skew that produces the industry's ROI, given the industry's cost. Unfortunately, simulations suggest skews that are too small to realistically represent pharmaceuticals, which illustrates friction: the presence of scant, incompatible data. To test a variety of skews, the simulations of Section 4 use a range spanning from low to high, $\delta \in \{1, 2, 4.18\}$, produced by lognormal distributions with $V \sim N(-0.049, 0.314^2)$, $V \sim N(-0.152, 0.551^2)$, and $V \sim N(-0.361, 0.85^2)$.

According to Figure 5, a skew of $\delta = 2$, with $k_D = P_{25}$, comes closest to the ROIs reported by Steedman et al.

These values help to check the simulation results for robustness by comparing the result of using lognormal distributions for X, with $\delta \in \{1, 2, 4.18\}$, to simulations that use exponential distributions. All exponential distributions have $\delta = 2$ and contain 33%, 19%, and 14% of their value in the top three deciles of the distribution, as do lognormal distributions with $\delta = 4.18$.

A2: Estimating the error in forecasts, ε

Fortunately, empirical data is more helpful for estimating σ_Y, which through equation (1) estimates ε. Cha et al. (2013) evaluated sell-side analysts' forecasts of peak sales, which is the maximum annual sales in a product's life-cycle, a key component in profit estimates. They calculated the relative error in forecasts of peak sales, $(\text{forecast} - \text{actual})/\text{actual}$, for each drug and then calculated the standard deviation of the results, creating a statistic analogous to σ_Y in the CRE model. Studying 260 drugs launched between 2002 to 2011, but excluding 54 of them as outliers, rejecting them for having forecasting errors exceeding 160%, Cha et al. calculated the standard deviation of the relative errors to be 75%. How can the error be so large, especially if underestimates are at most 100% too low? Overestimates may greatly exceed the true value, and because of the optimizer's curse, as Figure 2 shows, we expect these large overestimates. Including the "outliers" raises the estimated error to a minimum of $\varepsilon \geq 93\% = 79\% * 75\% + 21\% * 160\%$.

Three adjustments complete this calculation. The data from Cha et al. suffer from survivorship bias. They studied launched drugs, which usually have above-average profit

forecasts, with some of the forecasts being optimistic, thus inflating error estimates, which is another manifestation of the optimizer's curse. Hay et al. (2014) report that only 16.2% of phase 2 compounds achieve launch. For this throughput, Section 4's simulations estimate that survivorship bias inflates forecasting errors by 9%, so we should reduce the minimum error from 93% to 86%. Second, Cha et al. calculated the error in forecasts made two years before launch, a time when phase 3 trials are nearly complete. The advancement of compounds from phase 2 to phase 3 occurs, on average, four years before launch (Paul et al. 2010), and obviously, phase 3 data is unavailable for these decisions. For forecasting over a longer horizon and using less data, we should increase the error estimate. Third, peak sales is a critical component of profit forecasts, but other estimates contribute as well. Each estimate adds error to the forecast, suggesting a second upward revision. Considering the two upward revisions, $100\% \leq \sigma_Y \leq 125\%$ seems reasonable.

Using equation (1), Table 4 presents the forecast uncertainty for various skews of X. A skew of $\delta = 2$, coupled with high development costs, produces the ROIs close to those measured by Steedman (2018). However, the forecasting error for this skew, $\varepsilon > 79\%$, is quite high. A lower value from the table, $\varepsilon > 66\%$, maybe more realistic. Notice that ε decreases as δ increases (see Table A1) and recall the discussion of δ and ROI from above (Section A1). A skew of $\delta = 4.18$ is probably too high because, as Figure 5 reveals, it produces ROIs higher than those reported by Steedman et al. (2018). Thus, we can consider $\varepsilon > 66\%$ to be a lower bound on the forecasting error. Notice how little of the forecast is signal: $1 - \varepsilon < 34\%$. To test a variety of forecasting errors, spanning realistic values, from exceptional to terrible forecasts, the simulations test $\varepsilon \in \{20\%, 36\%, 50\%, 61\%, 70\%, 80\%\}$.

Skew of lognormal X	ε when $\sigma_Y = 100\%$	ε when $\sigma_Y = 125\%$
1.00	91%	94%
2.00	79%	86%
2.00 (exponential X)	67%	76%
4.18	66%	75%

Table A1: The forecasting uncertainty produced by $100\% \leq \sigma_Y \leq 125\%$ for various distributions of X, calculated with equation (1).

A3: Estimating technical uncertainty, AUC

Like the market parameters, empirical evidence guides the choice of technical parameters, π_E and AUC. Focusing on efficacy, statisticians sometimes size phase 2 trials to produce false-positive and false-negative rates of 5% and 20% (Chen et al. 2015; Lindborg et al. 2014), which from equation (4), produces a phase 2 resolution of $AUC_2 = 0.961$. The sizing of a clinical trial rests on uncertain assumptions, and if uncertainty is more harmful than helpful, on average, $AUC_2 = 0.961$ is an upper bound on phase 2's resolution. Uncertainty about safety reduces the resolution a well.

Lo et al. (2019) provide a lower bound. Using data from 4,812 phase 2 drugs, they developed classifiers to predict FDA approval. Drugs from the most recent period they studied, 2010-2014, produces the best classifiers, achieving $AUC_2 = 0.797$, with a 90% confidence interval of $[0.718, 0.876]$. For two reasons, phase 2 trials should produce resolution greater than these classifiers yield. First, while both phase 2 trials and Lo et al. predict success using characteristics of clinical trials, including results, clinical trials predict efficacy while Lo et al. predict FDA approval. Compounds canceled for nontechnical reasons, such as poor profitability or changes in corporate strategy, dull the resolution of Lo et al.'s classifiers. Second, while all approved drugs are effective, drugs that fail in development include both ineffective compounds (true-negatives) and effective ones (false-negatives). The placement of effective drugs among

both successes and failures dulls the differences between successful and failed compounds and reduces the classifiers' resolutions. In total, a reasonable range for technical resolution is $0.797 \leq AUC_2 \leq 0.961$ To test regimes of high and low resolutions, Section 4 uses $AUC_2 \in \{80\%, 90\%\}$.

A4: Estimating the percentage of projects that are technically sound, π_E

Estimating the fraction of phase 2 compounds that are effective, π_E, requires careful disentangling of intertwined concepts and equations. We start by estimating the fraction of phase 3 compounds that are effective and then use the results to analyze phase 2. Beginning with historical data from Hay et al. (2014), industry-wide, phase 3's attrition is 39.9%, of which 54% fail for efficacy issues, so $21.5\% = 39.9\% * 54\%$ of phase 3 compounds fail for demonstrating insufficient efficacy. If managers considered efficacy only, ignoring safety and business issues, 78.5% of phase 3 compounds would advance to a new drug application (NDA) with the FDA. About 16.8% of NDA applications fail, of which 48% percent fail for insufficiently demonstrating efficacy, so $92\% = 1 - 16.8\% * 48\%$ of NDA compounds are effective.

Plugging these values into equations (6) and (7), with $t_T = 78.5\%$ and $\pi_D = 92\%$, produces two equations with three unknowns: $1 - \Phi_S(v_c)$, $1 - \Phi_F(v_c)$, and π_3. From Section 2.2, we can rewrite $1 - \Phi_S(v_c)$, and $1 - \Phi_F(v_c)$ in equations (6) and (7) in terms of β and $\alpha/2$ and then use the relationship of β and $\alpha/2$ to AUC from equation (4). The resulting equations (4, 6, and 7) provide what we need. For any value of π_E, only one value of AUC produces β and $\alpha/2$ that satisfy equations (6) and (7). If we can get upper and lower bounds on AUC for phase 3, we can calculate the upper and lower bounds on π_3.

We find bounds on phase 3's AUC_3 using the same approach for finding AUC_2 (above). For the upper bound, statisticians typically size phase 3 trials to achieve a false-positive rate of 2.5% (one-tailed) with a false-negative rate of 20% (Senn 2007), which produces $AUC_3 = 0.976$ (equation 4). For the lower bound, Lo et al. (2019) produced classifiers for predicting approval with phase 3 results. The best classifiers they produced achieved $AUC_3 = 0.876$, with a 90% confidence interval of $[0.724, 1.000]$. Assuming $0.876 \leq AUC_3 \leq 0.976$, and using the historical estimates of $t_T = 78.5\%$ and $\pi_D = 92\%$, equations 4, 6 and 7 imply that the fraction of phase 3 compounds that are effective is $74\% \leq \pi_3 \leq 81\%$.

Now we repeat the calculations by applying equations 4, 6, and 7 to phase 2. In these equations, phase 2's throughput is t_T. The fraction of phase 2 compounds that are effective, the variable we need to estimate, is π_2 (replaces π_E in equations 6 and 7). The output of phase 2's selection, the fraction of phase 3 compounds that are effective, π_3, we just estimated (replaces π_D in equation 7). From historical data, 67.6% of phase 2 compounds fail (Hay et al. 2014), and 51% of these failures are for lacking efficacy (Arrowsmith 2011), so 34.5% of phase 2 compounds fail for being ineffective. Phase 2's (efficacy-only) throughput is $t_T = 65.5\%$. Using $74\% \leq \pi_3 \leq 81\%$ and $0.797 \leq AUC_3 \leq 0.961$ (estimated above), implies $49\% \leq \pi_2 \leq 66\%$. Some therapeutic areas, like oncology, have high failure rates, so to include regimes of fruitful and less fruitful choice sets, Section 4's simulations use $\pi_E \in \{35\%, 53\%\}$.

A5: Estimating development costs, k_D

Development cost is the final parameter to estimate. Recall that for convenience $E[X] = 1$ in all the simulations. With this mean, consider the impact of skew. Large skews spread X outward in both directions, while small skews make more compact curves. The latter case

produces some nonsensical results. With $\delta = 1$, a cost of 0.5, an absurdly high cost, sits below 98% of X, so only 2% of projects are unprofitable. To prevent these problems and represent costs consistently, we set k_D as a percentile of X. For concise notation, let P_a be the value of X at the a^{th} percentile, $k_D = P_a = F_X^{-1}(a)$.

Empirical data plus assumptions about δ estimate k_D. Grabowski et al. (2002) report that 34% of pharmaceuticals that achieve launch have profits exceeding the expected capitalized out-of-pocket costs (COC) of discovery and development. For a lognormal X with $E[X] = 1$ and $\delta = 4.18$, the 66^{th} percentile occurs at a cost of 0.99. (For convenience, assume random selection from X, rather than selection based on evaluations.) Grabowski et al. included post-launch expenditures in their calculations. Removing these expenses reduces the cost to 0.88. Meanwhile, DiMasi et al. (2003) report that phase 3 consumes 40% of the COC of creating a drug, while Paul et al. (2010) report that the FDA application consumes about 5% and that the success rates for phase 3 and the NDA are 70% and 91%, respectively. When selecting phase 2 compounds to advance to phase 3 trials (to advance to development in this paper's model), $\pi_D = 70\% * 91\% = 63.7\%$ and $k_D = (40\% + 5\% * 70\%) * 0.88 = 0.38$. As a percentile of X, the expected cost of a compound that advances to phase 3 is $k_D = P_{24.5}$. (Note: successful drugs must pay for the $1 - \pi_D$ projects that fail in phase 3, so a compound is profitable if $x > k_D/\pi_D = 0.61$, which occurs at $P_{43.6}$. Figure 3 uses this number.) Repeating these calculations for a lognormal X with a $\delta = 2$ produces a lower cost of $k_D = P_{10}$. To test both low and high costs, the simulations used $k_D \in \{P_5, P_{25}\}$.

A6: Calculating $E[X|Z=z]$ for the CRE model

For the CRE model, we calculate $E[X|Z=z]$ by deriving a formula for $f_{X|Z=z}(x)$ and then integrating $\int x f_{X|Z=z}(x)\,dx$ numerically. The formula for a conditional probability density for a continuous function is $f_{X|Z=z}(x) = f_{X,Z}(x,z)/f_Z(z)$. Starting with the denominator, $f_Z(z)$, the forecast in the CRE model is the product of two independent random variables, $Z = X + XY = X(1+Y)$. Springer (1979) derives the density function for the product of two independent random variables. We use the substitution $Y = (Z/X) - 1$ in his formula to get $f_Z(z) = \int (1/x) f_X(x) f_Y\big((z/x) - 1\big)\,dx$, integrated over $0 < x \leq \infty$.

To get the formula for $f_{X,Z}(x,z)$, note that the likelihood of a realization of $X = x$ and $Z = z$ is the likelihood of two events in a product distribution, $X = x$ and $Y = (z/x) - 1$, which is given by the integrand of $f_Z(z)$. Thus, $f_{X,Z}(x,z) = f_X(x) f_Y\big((z/x) - 1\big)/x$, for $x > 0$. Altogether, we have $f_{X|Z=z}(x) = f_X(x) f_Y\big((z/x) - 1\big)/[x f_Z(z)]$.

References

Al-Najjar NI, Pai MM (2014) Coarse decision making and overfitting. *J.Econom. Theory*. 150(1):467-486.

Arrowsmith, J (2011) Phase II failures: 2008-2010. *Nat. Rev. Drug Discovery*. 10(5):328.

Åsterbro T, Koehler DJ (2007) Calibration accuracy of a judgmental process that predicts the commercial success of new product ideas. *J. Behav. Decision Making*. 20(4):381-403.

Baker E, Bosetti V, Salo A. Robust portfolio decision analysis, College of Engineering, University of Massachusetts.

Beckman RA, Clark J, Chen C (2011) Integrating predictive biomarkers and classifiers into oncology clinical development programs. *Nat. Rev. Drug Discovery*. 10(10):735-748.

Berkeley D, Humphreys P (1982) Structuring decision problems and the 'bias heuristic.' *Acta Psychologica*. 50(3):201-252.

Bookstaber R, Langsam J (1985) On the optimality of coarse behavior rules. *J. Theoret. Biol*. 116(2): 161-193.

Brown HI (1994) Reason, judgment and Bayes's law. *Philos. Sci.* 61(3):351-369.

Carlin BP, Louis TA (2000) *Bayes and Empirical Bayes Methods for Data Analysis* (Chapman and Hall/CRC, New York).

Cha M, Rifia B, Sarraf P (2013) Pharmaceutical forecasting: throwing darts? *Nat. Rev. Drug Discovery*. 12(10):737-738.

Chaung-Stein C, Kirby S, French J, Kowalski K, Marshall S, Smith MK, Bycott P, Beltangady M (2011) A quantitative approach for making go/no-go decision in drug development. *Drug Information J.* 45(2):187-202.

Chen C, Beckman RA (2007) Optimal cost-effective designs of proof of concept trials and associated go-no go decisions. *ASA, Proceedings of the Biometrics Sect.*, 394-399.

Chen C, Beckman RA (2009a) Optimal cost-effective go-no go decision in late-stage oncology drug development. *Statist. Biopharmaceutical Res*. 1(2):159-169.

Chen C, Beckman RA (2009b) Optimal cost-effective designs of phase II proof of concept trials and associated go–no go decisions. *J. Biopharmaceutical Statist.* 19(3):424-436.

Chen C, Beckman RA, Sun LZ (2015). Maximizing return on investment in phase II proof-of-concept trials. Antonijevic Z, ed. *Optimization of Pharmaceutical R&D Programs and Portfolios: Design and Investment Strategy* (Springer, Cham, *Switzerland*), 141-154.

Chen M, Dyer J (2009) Inevitable disappointment in projects selected on the basis of forecasts. *SPE J*. 14(2):216-221.

Clemen RT (1989) Combining forecasts: A review and annotated bibliography. *Internat. J. Forecasting*. 5(4), 559-583

Clemen RT (2008) Improving and measuring the effectiveness of decision analysis: linking decision analysis and behavioral decision research. Kugler T, Smith JC, Conolly T, Son YJ, eds. *Decision Modeling and Behavior in Complex and Uncertain Environments* (Springer, New York), 3-31.

Clemen RT, Winkler RL (1985) Limits on the precision and value of information from dependent sources. *Oper. Res*. 33(2):427-442.

Cooper RG, Edgett SJ, Kleinschmidt, EJ (1998) *Portfolio Management for New Products* (Perseus Publishing, Cambridge, MA).

Cyert RM, March JG (1992) *A Behavioral Theory of the Firm* (Blackwell Publishers Inc., Malden, MA).

Dawes RM (1979) The robust beauty of improper linear models in decision-making. *Amer. Psych.* 34(7):571-582.

DeMiguel V, Garlappi L, Uppal R (2009) Optimal versus naive diversification: how inefficient is the $1/n$ portfolio strategy. *Rev. Financial Stud.* 22(5):1915-1953.

DiMasi, JA, Hansen RW, Grabowski HG (2003) The price of innovation: new estimates of drug development costs. *J. Health Econom.* 22(2):151-185.

Draper D (1995) Assessment and propagation of model uncertainty. *J. Roy. Statist. Soc. Ser. B Methodological*. 57(1): 45-98.

Einhorn HJ, Hograth RM (1981) Behavioral decision theory: processes of judgment and choice. *Annual Rev. Psych.* 32(1):53-88.

Elton EJ, Gruber MJ (1995) *Modern Portfolio Theory and Investment Analysis* (J Wiley, New York).

Fawcett T (2006) An introduction to ROC analysis. *Pattern Recognition Lett.* 27(8): 861-874.

Fliedner T, Liesiö J (2016) Adjustable robustness for multi-attribute project portfolio selection. *Eur. J. Oper. Res.* 252(3):931-946

French S (1995) Uncertainty and imprecision: modelling and analysis. *J. Oper. Res. Soc.* 46(1):70-79.

Frisch D, Clemen RT (1994) Beyond expected utility: rethinking behavioral decision research. *Psych. Bull.* 116(1):46-54.

Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB (2013) *Bayesian Data Analysis* (Chapman and Francis Group, Boca Raton, FL).

Gigerenzer, G (2008) Bounded and rational. *Rationality for Mortals: How People Cope with Uncertainty*. (Oxford University Press: Oxford), 3-19.

Gönen, M (2007) *Analyzing Receiver Operating Characteristic Curves with SAS* (SAS Institute Inc., Cary, NC).

Grabowski HG, Vernon J, DiMasi JA (2002) Returns on research and development for 1990s new drug introductions. *PharmacoEconom.* 18(Suppl. 1):21-32.

Harrison JR, March JG (1984) Decision making and postdecision surprises. *Admin. Sci. Quart*. 29(1):26-42.

Hassanzadeh F, Nemati H, Sun M (2014) Robust optimization for interactive multiobjective programming with imprecise information applied to R&D project portfolio selection. *Eur. J. Oper. Res*. 238(1):41-53.

Hay M, Thomas DW, Craighead JL, Economides C, Rosenthal J. (2014) Clinical development success rates for investigational drugs. *Nature Biotech.* 32(1):40-51.

Heidenberger K, Stummer C (1999) Research and development project selection and resource allocations: a review of quantitative modeling approaches. *Internat. J. Management Rev*. 1(2):197-224.

Heiner RA (1983) The origin of predictable behavior. *Amer. Econom. Rev*. 73(4):560-595.

Heiner RA (1986) Uncertainty, signal-detection experiments, and modeling behavior. Langlois R, ed. *Economics as a Process: Essays in the New Institutional Economics* (Cambridge University Press, New York), 59-115.

Heiner RA (1988) The necessity of imperfect decisions. *J. Econom. Behav. Organ*. 10(1):29-55.

Jones BD (1999) Bounded rationality. *Annual. Rev. Political Sci*. 2(1):297-321.

Jørgensen M (2013) The influence of selection bias on effort overruns in software development projects. *Inform. Software Tech*. 55(9):1640-1650.

Kahneman D, Lovallo D (1993) Timid choices and bold forecasts: a cognitive perspective on risk taking. *Management Sci*. 39(1):17-31.

Kahneman D, Tversky A (1982) Intuitive prediction: biases and corrective procedures. Kahneman D, Slovic P, Tversky A, eds. *Judgment Under Uncertainty: Heuristics and Biases* (Cambridge University Press: Cambridge), 414-421.

Kavadias S, Chao RO (2008) Resource allocation and new product development portfolio management. Loch CH, Kavadias S, eds. *Handbook of New Product Development Management* (Elsevier: Amsterdam), 135-163.

Keisler J (2004) The value of information in portfolio decision analysis. *Decision Anal*. 1(3):177-189.

Kettunen J, Salo A (2017) Estimation of downside risks in project portfolio selection. *Production Oper. Management*. 26(10):1839-1853.

Keeney, RL (1992) *Value-Focused Thinking: A Path to Creative Decisionmaking* (Harvard University Press: Cambridge, MA).

Kloeber J (2011) Current and cutting edge methods of portfolio decision analysis in pharmaceutical R&D. Salo A, Keisler J, Morton A, eds. *Portfolio Decision Analysis: Improved methods for resource allocation* (Springer, New York), 283-331.

Lendrem DW, Lendrem BC (2013) Torching the haystack: modeling fast-fail strategies in drug development. *Drug Discovery Today*. 18(7/8):331-336.

Lenk PJ, Rao AJ (1990) New models from old: forecasting product adoption by hierarchical Bayes procedures. *Marketing Sci*. 9(1):42-53.

Liesiö J, Mild P, Salo A (2007) Preference programming for robust portfolio modeling and project selection. *Eur. J. Oper. Res.* 181(3):1488-1505.

Liesiö J, Mild P, Salo A (2008) Robust portfolio modeling with incomplete cost information and project interdependencies. *Eur. J. Oper. Res*. 190(3):679-695.

Liesiö, J, Salo A (2012) Scenario-based portfolio selection of investment projects with incomplete probability and utility information. *Eur. J. Oper. Res.* 217(1):162-172.

Lindborg SR, Persinger CC, Sashegyi A, Mallinckrodt C, Ruberg SJ (2014) Statistical refocusing in the design of phase II trials offers promise of increased R&D productivity. *Nat. Rev. Drug Discovery*. 13(8):638-640.

Lindley DV (1986) The reconciliation of decision analysis. *Oper. Res*. 34(2):289-295.

Lo AW, Siah KW, Wong CH (2019) Machine learning with statistical imputation for predicting drug approvals. Working paper, Sloan School of Management, Massachusetts Institute of Technology, Cambridge, Massachusetts, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2973611.

Lovallo D, Kahneman D (2003) Delusions of success – how optimism undermines executives' decisions. *Harvard Bus. Rev*. 81(7):56-63.

Macmillan NA, Creelman CD (2005) *Detection Theory: A User's Guide* (Lawrence Erlbaum Associates: Mahwah, NJ).

March JG, Sproull LS, Tamuz M (1991) Learning from samples of one or fewer. *Organ. Sci.* 2(1): 1-11.

Neelamegham R, Chintagunta P (1999) A Bayesian model to forecast new product performance in domestic and international markets. *Marketing Sci*. 18(2):115-136.

Michaud RO (1989) The Markowitz optimization enigma: is 'optimized' optimal? *Financial Analysts J*. 45(1):31-42.

Michaud RO (1998) *Efficient Asset Allocation: A Practical Guide to Stock Portfolio Optimization and Asset Allocation* (Harvard Business School Press, Boston).

Montibeller G, von Winterfeldt D (2015) Cognitive and motivational biases in decision and risk analysis. *Risk Analysis*. 35(7):1230-1251.

Pammer SE, Fong DKH, Arnold SF (2000) Forecasting the penetration of a new product – a Bayesian approach. *J. Bus. Econom. Statist.* 18(4):428-435.

Patel NR, Ankolekar S. (2007) A Bayesian approach for incorporating economic factors in sample size design for clinical trials of individual drugs and portfolios of drugs. *Statist. Med*. 26(27):4976-4988.

Patel NR, Ankolekar S (2015) Dynamically optimizing budget allocation for phase 3 drug development portfolios incorporating uncertainty in the pipeline. Antonijevic Z, ed. *Optimization of Pharmaceutical Programs and Portfolios: Design and Investment Strategy* (Springer International Publishing: Switzerland), 181-200.

Paul SM, Mytelka DS, Dunwiddie CT, Persinger CC, Munos BH, Lindborg SR, Schacht AL (2010) How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat. Rev. Drug Discov*. 9(3):203-214.

Ravinder HV, Kleinmuntz DN, Dyer JS (1988) The reliability of subjective probabilities obtained through decomposition. *Management Sci.* 34(2):186-199.

Salo A, Keisler J, Morton A (2011) *Portfolio Decision Analysis: Improved methods for resource allocation* (Springer, New York).

Savage LJ (1954) *The Foundations of Statistics* (Wiley, New York).

Schuyler J, Nieman T (2008) Optimizer's curse: removing the effect of this bias in portfolio planning. *SPE Projects Facilities Construction*. 3(1):1-9.

Senn S (2007) *Statistical Issues in Drug Development* (John Wiley and Sons, Chichester, England).

Simon, HA (1976) From substantive to procedural rationality.S. Latsis, ed. *Method and Appraisal in Economics* (Cambridge University Press, Cambridge), 65-86.

Slovic P, Fischhoff B, Lichtenstein S (1977) Behavioral decision theory. *Annual Rev. Psych.* 28(1):1-39.

Smith JE, Winkler RL (2006) The optimizer's curse: skepticism and postdecision surprise in decision analysis. *Management Sci*. 52(3):311-322.

Spetzler C, Hannah W, Mayer J (2016) *Decision Quality: Value Creation from Better Business Decisions* (Wiley, Hoboken, NJ).

Steedman M, Taylor K, Stockbridge M, Korba C, Shah Sonal, Thaxter M (2018) Unlocking R&D productivity: measuring the return from pharmaceutical innovation 2018. Deloitte Centre for Health Solutions, Deloitte LLP Life Sciences and Health Care Practices.

Stonebraker JS, Keisler J (2011) Empirically investigating the portfolio management process: findings from a large pharmaceutical company. Salo A, Keisler J, Morton A, eds. *Portfolio Decision Analysis: Improved methods for resource allocation* (Springer, New York), 131-148.

Springer MD (1979) *The Algebra of Random Variables* (Wiley: New York).

Sultan F, Farley JU, Lehmann DR (1990) A meta-analysis of applications of diffusion models. *J. Marketing Res.* 27(1):70-77.

Summers GJ (2010) Identify projects with the greatest chance of success. *BioPharmaPM Newsletter*. 12: 12-20.

Thaler RH (1988) Anomalies: the winner's curse. *J Econom. Perspectives*. 2(1):191-202.

Villumaa E. Liesiö J, Salo A (2014) Optimal Strategies for selecting project portfolios using uncertainty value estimates. *Eur. J. Oper. Res*. 233(3):772-783.

Vilkkumaa E, Salo A, Liesiö J, Siddiqui A (2015) Fostering breakthrough technologies – How do optimal funding decisions depend on evaluation accuracy? *Tech. Forecasting Soc. Change*. 96:173-190.

von Neumann J, Morgenstern O (1953) *Theory of Games and Economic Behavior* (Princeton University Press, Princeton).

Williamson OE (1989) Transaction cost economics. Schmalensee R, Willig RD, eds. *Handbook of Industrial Organization, Volume I* (Elsevier Science, Amsterdam), 135-182.

Zan K, Bickel E (2013) Components of portfolio value of information. *Decision Anal*. 10(2):171-185.