The Optimizer’s Curse

How uncertainty + project selection create mediocre, overvalued portfolios;

how Risk Optimizer can fail;

&

how to mitigate these problems.
Structural Uncertainty is Profound

**Study:** 29 teams of scholars analyze data from European soccer leagues to answer the following question: “Are darker skinned players more likely to receive red cards?”

<table>
<thead>
<tr>
<th>Technique</th>
<th># of model variations (flavors)</th>
<th># of methods for correlated errors</th>
<th># of covariates</th>
<th>Range of odds ratios</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
<td>6</td>
<td>3</td>
<td>3-7</td>
<td>[1.03, 1.28]</td>
</tr>
<tr>
<td>Logistic</td>
<td>15</td>
<td>3</td>
<td>0-6</td>
<td>[0.96, 1.48]</td>
</tr>
<tr>
<td>Poisson</td>
<td>6</td>
<td>3</td>
<td>1-6</td>
<td>[0.89, 2.93]</td>
</tr>
<tr>
<td>Other</td>
<td>2</td>
<td>2</td>
<td>0-3</td>
<td>[1.71, 2.88]</td>
</tr>
</tbody>
</table>

**Modeling building process:** Teams discuss the modeling with each other, catch and fix each other’s errors, ensuring all models are technically sound, exercising good judgment for every decision.

**Result:** 29 different models, each one correct, valid, and defensible, producing a wide variety of results. **Structural uncertainty is profound.**
Empirical Data Reveals Profound Uncertainty

Real data, estimating the average (standard deviation) of forecasting errors of peaks, a key compound of revenue forecasts.

- Two years prior to launch, the END of development (-2, on the horizontal axis) average forecasting errors are 75% of the true value.
- Actual errors are worse because selection occurs four years prior to launch for phase 3 drugs and 6.5 years before launch for phase 2 drugs.
- The study’s authors, from McKinsey, eliminated 20% of the data. Including this data, data the optimizer’s curse predicts, raises the forecasting two year before launch to at least 92%.
Inaccuracy vs. imprecision

- Inaccurate (bias) but precise
- Imprecise but accurate (unbiased)

- Decision analysis worries about biased data.
- Imprecision causes the optimizer’s curse
- Impact of imprecision on selection is terrible (next slide)
Using @Risk, you can perform this simulation experiment

**Step 1:** Create 1000’s of randomly generated true value for projects. To do this, ask, “How are profits distributed in my industry?” Most industries exhibit the Pareto rule in which small percent of products produce most of the profits. In pharma, 30% of products produce 80% of the industry’s profits. Generate the “True Value” column (below) with an asymmetric distribution, such as a lognormal distribution.

**Step 2:** Randomly generate unbiased forecasting errors using a normal distribution with standard deviation that represents the forecasting errors in your industry (for pharma, see slide 3).

**Step 3:** Combine the true values \(X\) with the errors \(Y\) to create forecasts \(Z\). Use these formulas: (a) most industries: \(Z = X + XY\) or (b) Oil & Gas: \(Z = e^{X+Y}\).

You now have thousands of true values and forecasts that accurately. With them, we can learn much. See next slide.
Distribution of Profits (NPV; True Value; $X$)

1 = average profit for a product in your industry
2 = twice the average profit
4 = four times the average profit
$Z = 4$ means 4 times the average profit – a good definition of a blockbuster
Question: If the forecast is $Z = 4$, what is the expected true value?

If the forecast is $Z = 4$, which is a blockbuster, what should we think about the true value?

Hypothesis: If $Z = 4$, the $E[X] = 4$. Here is the reasoning:

A. A forecast is equally likely to be above or below the true value
B. The true value is equally likely to be above or below the forecast
C. Most graduate students eat Ramen noodles
D. Most people who eat Ramen noodles are graduate students
On average, above average forecasts are too high

Rare disease affects 1 out 100 people. Diagnostic test has the following qualities:
- Always identifies people who have the disease
- False-positive rate is 5%

Test 100 people and get 6 positive results: 1 true-positive and 5 false-positives
A positive test has a means a 17% of having the disease.

Many opportunities to overestimate the value of these projects
Few opportunities to overestimate the value of these projects

Forecast (z) in units of average market size
Average over- or Underestimate
0 1 2 3 4 5 6
-80% -40% 0 20% 40% 60% 80% 100% 120%
Impact of the Optimizer’s Curse on Simulation Optimization (Risk Optimizer)

Setting Up the Model

For twenty-five projects:
1. Randomly generate a true value with LogNormal (slide 8, left distribution)
2. Randomly generate a forecasting error (slide 8, right distribution)
3. Combine to create a forecast, \( Z = X + XY \)
4. Using a cutoff value, select projects based on the forecast
5. Via Monte Carlo analysis, estimate the portfolio’s:
   - mean and standard deviation predicted by the projects’ forecasts, \( Z \)
   - true mean and standard deviation, given by the projects’ true values, \( Z \)
6. For oil & gas companies, repeat the tests with a lognormal forecast model, \( Z = e^{X+Y} \)
100% of Projects Selected

- Forecasted portfolio value (red) & True portfolio value (blue)
- No optimizer’s curse

Forecast Overestimates
Expected value: 0%
Standard Deviation: 82%

Oil & Gas

Most Industries
Illustrating the Optimizer’s Curse
Forecasted portfolio value (red) & True portfolio value (blue)
90% of projects selected (bottom 10% rejected)

Forecast Overestimates
Expected value: 15%
Standard Deviation: 62%

Forecast Overestimates
Expected value: 49%
Standard Deviation: 184%
Illustrating the Optimizer’s Curse
Forecasted portfolio value (red) & True portfolio value (blue)
50% of projects selected (bottom 50% rejected)

Forecast Overestimates
Expected value: 42%
Standard Deviation: 52%

Forecast Overestimates
Expected value: 94%
Standard Deviation: 156%
ROI of Phase III Portfolios is Below the Cost of Capital (10%)

• No one would fund projects with ROIs below the cost of capital, so how can large, diversified portfolios consistently underperform?

• If we build good models, fill them with unbiased data, perform sensitivity and Monte Carlo analysis with @Risk, doing everything correctly, as large pharmaceutical companies do, why do their portfolios persistently underperform?

• For each project selection method – cutoff values, hurdle rates, portfolio optimization, both simple and sophisticated, and simulation optimization (Risk Optimizer in @Risk):
  • In what situations does each method fail?
  • Why do they fail?
  • What does failure look like? What are the consequences?
  • How can we keep them from failing?

Impact of the Optimizer’s Curse on Forecasts of Portfolio Mean and Standard Deviation

These incorrect estimates cause simulation optimization (Risk Optimizer) to:

- Recommend substantially suboptimal portfolios
- Produce incorrect expectations of portfolio value and risk
- Adding details (constraints on risk, resource constraints, project interactions, scheduling) exacerbates these problems
Using Monte Carlo analysis, perhaps applied to these spreadsheet models, if we estimate projects’ values with probability distributions, instead of point estimates, do we eliminate the optimizer’s curse?

You can mitigate the optimizer’s curse when using either point estimates or probability distributions, but the fix is counter-intuitive. When using point estimates the fix is easier to apply. Moreover, without the fix, Risk Optimizer is a disaster for portfolio optimization. All applications that I have seen make the same error.
The Common Error Made with Risk Optimizer

Suppose the Monte Carlo simulation produces this curve. The mean is $Z = 4$, and there is a probability distribution, perhaps symmetrical, perhaps not, around the mean. The mean is still too high (see chart to the right), and the probability distribution is wrong too: too high, wrong shape, with the wrong standard deviation.

A. The correct point estimates, or correct probability distributions require data, additional data, from outside of your model, from outside your Monte Carlo analysis. It requires additional data.

B. This is easy to understand. A forecast of great success, such as $Z = 4$, is on average too high. We cannot fix this with the forecast because we have already used that data. That data is what created the, likely, overestimate. We need new, additional data.

Forecast (z) in units of average market size

Average over- or Underestimate
How to Mitigate the Optimizer’s Curse

Distribution of profits for class of products
Class data
Outside view
Bayesian prior distribution

Your forecast: Monte Carlo simulation of your forecasting model
Case data
Inside view
New information to combine with the prior

Combine both means (point estimates) or probability distributions using either:

A. Bayes’ law
B. Poor man’s Bayes law (requires less data, great for point estimates)

Create new point estimate or new probability distribution
How to Succeed at Portfolio Management

A. Portfolio management confronts a tremendous amount of structural and parametric uncertainty.

B. Success comes not from optimizing or explicitly try to maximize portfolio value.

C. Success comes from managing uncertainty well, from mitigating the damage that uncertainty causes.

D. Minimizing error creates more success than explicitly striving to maximize value.

Managing Uncertainty in Portfolio Management

1. Better choice sets
   a) Overall quality
   b) Shots on goal

2. Better forecasts
   a) Simplicity v. sophistication
   b) Empirical data
   c) Averaging forecast
   d) Mitigating the optimizer’s curse

3. Robust selection techniques & metrics
   a) Non-compensatory metrics & screens
   b) Rankings (ROI, eNPV)
   c) Simple optimization
   d) Complex optimization
   e) Simulation optimization

4. Agile portfolio management

5. Learn from result: improve with every iteration
The Optimizer’s Curse

How uncertainty + project selection create mediocre, overvalued portfolios;

how Risk Optimizer can fail;

&

how to mitigate these problems.