Additional optional reading


1.1–3, 2.4–6, 3.1, 3.4, 4, 5.1.
Motivation

• Until now we assumed to know all the *parameters* of the problem exactly.
Motivation

- Until now we assumed to know all the parameters of the problem exactly. E.g., in

  \[
  \begin{align*}
  \min & \quad c^T x \\
  \text{s.t.} & \quad Ax = b \\
  & \quad x > 0
  \end{align*}
  \]

  we assume to know the exact values for $A$, $b$, and $c$. 
Motivation

• Until now we assumed to know all the *parameters* of the problem exactly. E.g., in

\[
\begin{align*}
\text{min } & c^T x \\
\text{s.t. } & Ax = b \\
& x > 0
\end{align*}
\]

we assume to know the *exact values* for \( A, b, \) and \( c \).

• What if \( A \) and/or \( b \), and/or \( c \) were *random variables with known distribution*?
Uncertainty

- $\Omega = \{\omega_1, \ldots, \omega_S\}$: sample space (partitioned into $S$ disjoint events)
- $p_i = \Pr(\omega_i)$: probability of event $\omega_i$, $\sum_{i=1}^{S} p_i = 1$.
- $\omega$: random parameters: i.e., random variable (or vector) representing events in $\Omega$;
- $A = A(\omega), b = b(\omega), c = c(\omega)$
Uncertainty

• $\Omega = \{\omega_1, \ldots, \omega_S\}$: sample space (partitioned into $S$ disjoint events)

• $p_i = \Pr(\omega_i)$: probability of event $\omega_i$, $\sum_{i=1}^{S} p_i = 1$.

• $\omega$: random parameters: i.e., random variable (or vector) representing events in $\Omega$;

• $A = A(\omega), b = b(\omega), c = c(\omega)$

*Stochastic programming* is the area of mathematical optimization dealing with uncertainties.
Uncertainty

- $\Omega = \{\omega_1, \ldots, \omega_S\}$: sample space (partitioned into $S$ disjoint events)
- $p_i = \Pr(\omega_i)$: probability of event $\omega_i$, $\sum_{i=1}^{S} p_i = 1$.
- $\omega$: random parameters: i.e., random variable (or vector) representing events in $\Omega$;
- $A = A(\omega)$, $b = b(\omega)$, $c = c(\omega)$

Stochastic programming is the area of mathematical optimization dealing with uncertainties.

**Question**

What about the decision variables $x$? Do they depend on $\omega$ or not?
Anticipative and adaptive decisions

Stochastic programming models can have two kinds of decision variables:

- **anticipative** (or first stage): decisions that must be taken before the realization of the random parameter $\omega$ is known;
- **adaptive** (or corrective, or second stage): decisions that can be made after some (or all) the random parameters have been revealed.
Example: The uncertain life of a European farmer

- Mario, a European farmer, has 500 acres of land;
- He specializes in wheat, corn, and sugar beets;
- Mario knows that the mean yield on his land is:
  - 2.5 tons (T) of wheat per acre
  - 3 T of corn per acre
  - 20 T of sugar beets per acre
- During the winter he must decide how much land to devote to each crop;
Example: The uncertain life of a European farmer

- Mario, a European farmer, has 500 acres of land;
- He specializes in wheat, corn, and sugar beets;
- Mario knows that the mean yield on his land is:
  - 2.5 tons (T) of wheat per acre
  - 3 T of corn per acre
  - 20 T of sugar beets per acre
- During the winter he must decide how much land do devote to each crop;
- Planting crops incurs in costs;
- He has some constraints on the production (see later);
- If he satisfies the constraints, he can sell excess production, otherwise he must buy what he needs;
- Mario wants to minimize the losses (difference between expenses for planting and buying and profits for selling)
Wheat and corn:
• Mario needs 200 T of wheat and 240 T of corn are needed for cattle feed;
• Can be raised on the farm, or bought from a wholesaler;
• Any production in excess would be sold;
• Mean selling prices: $170 per T of wheat, $150 per T of corn;
• Buying prices: $238 and $210 (40% more than selling);
• Planting costs: $150 per acre of wheat, $230 per acre of corn;

Sugar beets:
• Mario has no needs for beets, but they are profitable.
• The European Commission imposes quotas on beets production.
• Mario’s quota is 6000 T;
• He can sell up to 6000 T for $36 per T, and anything in excess for $10 per T;
• Planting beets costs $260 per acre;
Constraints, costs, and profits

Wheat and corn:
- Mario needs 200 T of wheat and 240 T of corn are needed for cattle feed;
- Can be raised on the farm, or bought from a wholesaler;
- Any production in excess would be sold;
- Mean selling prices: $170 per T of wheat, $150 per T of corn;
- Buying prices: $238 and $210 (40% more than selling);
- Planting costs: $150 per acre of wheat, $230 per acre of corn;

Sugar beets:
- Mario has no needs for beets, but they are profitable.
- The European Commission imposes quotas on beets production. Mario’s quota is 6000 T;
- He can sell up to 6000 T for $36 per T, and anything in excess for $10 per T;
- Planting beets costs $260 per acre;
## Data and variables

<table>
<thead>
<tr>
<th></th>
<th>Wheat</th>
<th>Corn</th>
<th>Sugar Beets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yield (T/acre)</td>
<td>2.5</td>
<td>3</td>
<td>20</td>
</tr>
<tr>
<td>Planting cost ($/acre)</td>
<td>150</td>
<td>230</td>
<td>260</td>
</tr>
<tr>
<td>Selling price ($/T)</td>
<td>170</td>
<td>150</td>
<td>36 under 6000 T</td>
</tr>
<tr>
<td>Purchase price ($/T)</td>
<td>238</td>
<td>210</td>
<td>10 above 6000 T</td>
</tr>
<tr>
<td>Minimum requirement (T)</td>
<td>200</td>
<td>240</td>
<td>–</td>
</tr>
</tbody>
</table>

Total available land: 500 acres
### Data and variables

<table>
<thead>
<tr>
<th></th>
<th>Wheat</th>
<th>Corn</th>
<th>Sugar Beets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yield (T/acre)</td>
<td>2.5</td>
<td>3</td>
<td>20</td>
</tr>
<tr>
<td>Planting cost ($/acre)</td>
<td>150</td>
<td>230</td>
<td>260</td>
</tr>
<tr>
<td>Selling price ($/T)</td>
<td>170</td>
<td>150</td>
<td>36 under 6000 T</td>
</tr>
<tr>
<td>Purchase price ($/T)</td>
<td>238</td>
<td>210</td>
<td>10 above 6000 T</td>
</tr>
<tr>
<td>Minimum require-</td>
<td>200</td>
<td>240</td>
<td>–</td>
</tr>
<tr>
<td>ment (T)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Total available land: 500 acres

- $x_1, x_2, x_3$: acres of land devoted wheat, corn, and beets
- $w_1, w_2, w_3, w_4$: sold T of wheat, corn, beets (at favorable price), beets (at unfavorable price);
- $y_1, y_2$: purchased T of wheat, corn;

The problem reads as follows:

$$
\begin{align*}
\text{min } & \quad 150 x_1 + 230 x_2 + 260 x_3 + 238 y_1 - 170 w_1 + 210 y_2 - 150 w_2 - 36 w_3 - 10 w_4 \\
\text{s.t. } & \quad x_1 + x_2 + x_3 \leq 500, \\
& \quad 5 x_1 + y_1 - w_1 \geq 200, \\
& \quad 3 x_2 + y_2 - w_2 \geq 240, \\
& \quad w_3 + w_4 \leq 20 x_3, \\
& \quad w_3 \leq 6000, \\
& \quad x_1, x_2, x_3, y_1, y_2, w_1, w_2, w_3, w_4 \geq 0.
\end{align*}
$$
Data and variables

<table>
<thead>
<tr>
<th></th>
<th>Wheat</th>
<th>Corn</th>
<th>Sugar Beets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yield (T/acre)</td>
<td>2.5</td>
<td>3</td>
<td>20</td>
</tr>
<tr>
<td>Planting cost ($/acre)</td>
<td>150</td>
<td>230</td>
<td>260</td>
</tr>
<tr>
<td>Selling price ($/T)</td>
<td>170</td>
<td>150</td>
<td>36 under 6000 T</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>10 above 6000 T</td>
</tr>
<tr>
<td>Purchase price ($/T)</td>
<td>238</td>
<td>210</td>
<td>–</td>
</tr>
<tr>
<td>Minimum requirement (T)</td>
<td>200</td>
<td>240</td>
<td>–</td>
</tr>
<tr>
<td>Total available land: 500 acres</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- $x_1, x_2, x_3$: acres of land devoted wheat, corn, and beets
- $w_1, w_2, w_3, w_4$: sold T of wheat, corn, beets (at favorable price), beets (at unfavorable price);
- $y_1, y_2$: purchased T of wheat, corn;

Formulation:

$$\begin{align*}
\min & \quad 150x_1 + 230x_2 + 260x_3 + 238y_1 - 170w_1 + 210y_2 \\
& - 150w_2 - 36w_3 - 10w_4 \\
\text{s.t.} & \quad x_1 + x_2 + x_3 \leq 500 \\
& \quad 2.5x_1 + y_1 - w_1 \geq 200, 3x_2 + y_2 - w_2 \geq 240 \\
& \quad w_3 + w_4 \leq 20x_3, w_3 \leq 6000 \\
& \quad x_1, x_2, x_3, y_1, y_2, w_1, w_2, w_3, w_4 \geq 0
\end{align*}$$
Formulation

\[
\begin{align*}
\text{min } & \quad 150x_1 + 230x_2 + 260x_3 + 238y_1 - 170w_1 + 210y_2 \\
& \quad - 150w_2 - 36w_3 - 10w_4 \\
\text{s.t. } & \quad x_1 + x_2 + x_3 \leq 500 \\
& \quad 2.5x_1 + y_1 - w_1 \geq 200, 3x_2 + y_2 - w_2 \geq 240 \\
& \quad w_3 + w_4 \leq 20x_3, w_3 \leq 6000 \\
& \quad x_1, x_2, x_3, y_1, y_2, w_1, w_2, w_3, w_4 \geq 0
\end{align*}
\]

Easy linear program. Solution:
min $150x_1 + 230x_2 + 260x_3 + 238y_1 - 170w_1 + 210y_2$
$- 150w_2 - 36w_3 - 10w_4$

s.t. $x_1 + x_2 + x_3 \leq 500$
$2.5x_1 + y_1 - w_1 \geq 200, 3x_2 + y_2 - w_2 \geq 240$
$w_3 + w_4 \leq 20x_3, w_3 \leq 6000$
$x_1, x_2, x_3, y_1, y_2, w_1, w_2, w_3, w_4 \geq 0$

Easy linear program. Solution:

<table>
<thead>
<tr>
<th>Culture</th>
<th>Wheat</th>
<th>Corn</th>
<th>Sugar Beets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surface (acres)</td>
<td>120</td>
<td>80</td>
<td>300</td>
</tr>
<tr>
<td>Yield (T)</td>
<td>300</td>
<td>240</td>
<td>6000</td>
</tr>
<tr>
<td>Sales (T)</td>
<td>100</td>
<td>–</td>
<td>6000</td>
</tr>
<tr>
<td>Purchase (T)</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Overall profit</td>
<td>$118,600</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The worried farmer

- After looking at the solution, Mario is worried;
- Over the years, yields of each crop have been very different, mainly due to weather conditions;
The worried farmer

• After looking at the solution, Mario is worried;
• Over the years, yields of each crop have been very different, mainly due to weather conditions;

Simplifying assumptions:
• years are good, fair, or bad for all crops:
  • good year: yield is 20% above mean;
  • fair year: yield is exactly mean;
  • bad year: yield is 20% below mean;
The worried farmer

After looking at the solution, Mario is worried;

Over the years, yields of each crop have been very different, mainly due to weather conditions;

Simplifying assumptions:

- years are good, fair, or bad for all crops:
  - good year: yield is 20% above mean;
  - fair year: yield is exactly mean;
  - bad year: yield is 20% below mean;

- prices are not influenced by weather;

Mario worries about what would happen in a good year and in a bad year
Good year / Bad year

Mario adjusts the yields in the LP formulation and solves two LP, one for good years, and one for bad years;

<table>
<thead>
<tr>
<th>Culture</th>
<th>Wheat</th>
<th>Corn</th>
<th>Sugar Beets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surface (acres)</td>
<td>183.33</td>
<td>66.67</td>
<td>250</td>
</tr>
<tr>
<td>Yield (T)</td>
<td>550</td>
<td>240</td>
<td>6000</td>
</tr>
<tr>
<td>Sales (T)</td>
<td>350</td>
<td>–</td>
<td>6000</td>
</tr>
<tr>
<td>Purchase (T)</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Overall profit:</td>
<td>$167,667</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure: Good years
Good year / Bad year

Mario adjusts the yields in the LP formulation and solves two LP, one for good years, and one for bad years:

### Good years

<table>
<thead>
<tr>
<th>Culture</th>
<th>Wheat</th>
<th>Corn</th>
<th>Sugar Beets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surface (acres)</td>
<td>183.33</td>
<td>66.67</td>
<td>250</td>
</tr>
<tr>
<td>Yield (T)</td>
<td>550</td>
<td>240</td>
<td>6000</td>
</tr>
<tr>
<td>Sales (T)</td>
<td>350</td>
<td>–</td>
<td>6000</td>
</tr>
<tr>
<td>Purchase (T)</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Overall profit: $167,667</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure: Good years**

<table>
<thead>
<tr>
<th>Culture</th>
<th>Wheat</th>
<th>Corn</th>
<th>Sugar Beets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surface (acres)</td>
<td>100</td>
<td>25</td>
<td>375</td>
</tr>
<tr>
<td>Yield (T)</td>
<td>200</td>
<td>60</td>
<td>6000</td>
</tr>
<tr>
<td>Sales (T)</td>
<td>–</td>
<td>–</td>
<td>6000</td>
</tr>
<tr>
<td>Purchase (T)</td>
<td>–</td>
<td>180</td>
<td>–</td>
</tr>
<tr>
<td>Overall profit: $59,950</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure: Bad years**
Mario adjusts the yields in the LP formulation and solves two LP, one for good years, and one for bad years:

The optimal solution is very sensitive to changes in yields, and so is the profit.

(Profit for fair years: $118,600)
• Mario is even more worried now: he is unable to make a perfect decision that would be best in all circumstances;
• He would like to understand the benefits and losses of each decision in each situation;

E.g. $w_{32}$: T of beets sold at the favorable price if yields are average.
• Mario is even more worried now: he is unable to make a perfect
decision that would be best in all circumstances;
• He would like to understand the benefits and losses of each
decision in each situation;

• The decisions \( x_1, x_2, x_3 \) on land assignments must be done now
  known as first stage or anticipative variables
• But sales and purchases \((w_i, y_i)\) depends on yields
  adaptive or second-stage variables
• Mario is even more worried now: he is unable to make a perfect decision that would be best in all circumstances;
• He would like to understand the benefits and losses of each decision in each situation;

• The decisions \( x_1, x_2, x_3 \) on land assignments must be done now known as \emph{first stage} or \emph{anticipative} variables
• But sales and purchases \((w_i, y_i)\) depends on yields \emph{adaptive} or \emph{second-stage} variables

Let’s give the sales and purchases variables an additional subscript index, denoting the \emph{scenario} \( s \)

\((s = 1: \text{good year}, \ s = 2: \text{fair year}, \ s = 3: \text{bad year})\):
Mario is even more worried now: he is unable to make a perfect decision that would be best in all circumstances;

He would like to understand the benefits and losses of each decision in each situation;

The decisions \( x_1, x_2, x_3 \) on land assignments must be done now known as first stage or anticipative variables

But sales and purchases \((w_i, y_i)\) depends on yields adaptive or second-stage variables

Let’s give the sales and purchases variables an additional subscript index, denoting the scenario \( s \)

\[ (s = 1: \text{good year}, s = 2: \text{fair year}, s = 3 \text{ bad year}): \]

\( w_{is}: \ i = 1, 2, 3, 4, \ s = 1, 2, 3 \)

\( y_{is}: \ i = 1, 2, \ s = 1, 2, 3 \)

E.g.: \( w_{32} \): T of beets sold at the favorable price if yields are average.
Mario wants to maximize his long-run profit, i.e., his expected profit
(He’s risk-neutral)
Mario wants to maximize his long-run profit, i.e., his expected profit.
(He’s risk-neutral)

We need a probability distribution:
  good, fair, and bad years are equiprobable (each has prob. 1/3)
Mario wants to maximize his long-run profit, i.e., his expected profit (He’s risk-neutral)

We need a probability distribution:
good, fair, and bad years are equiprobable (each has prob. 1/3)

Mario’s LP then becomes:

\[
\begin{align*}
\text{min} & \quad 150x_1 + 230x_2 + 260x_3 \\
& - \frac{1}{3}(170w_{11} - 238y_{11} + 150w_{21} - 210y_{21} + 36w_{31} + 10w_{41}) \\
& - \frac{1}{3}(170w_{12} - 238y_{12} + 150w_{22} - 210y_{22} + 36w_{32} + 10w_{42}) \\
& - \frac{1}{3}(170w_{13} - 238y_{13} + 150w_{23} - 210y_{23} + 36w_{33} + 10w_{43}) \\
\text{s.t.} & \quad x_1 + x_2 + x_3 \leq 500, \\
& \quad 3x_1 + y_{11} - w_{11} \geq 200, \\
& \quad 3.6x_2 + y_{21} - w_{21} \geq 240, \\
& \quad 2.5x_1 + y_{12} - w_{12} \geq 200, \\
& \quad w_{32} + w_{42} \leq 20x_3, \\
& \quad w_{33} \leq 6000, \\
& \quad w_{33} + w_{43} \leq 16x_3, \\
& \quad w_{33} \leq 6000, \quad x, y, w \geq 0.
\end{align*}
\]

The optimal solution can be understood as follows. The most profitable decision for sugar beet land allocation is the one that always avoids sales at the unfavorable price even if this implies that some portion of the quota is unused when yields are average or below average.
### Solution

<table>
<thead>
<tr>
<th>First Stage</th>
<th>Wheat (acres)</th>
<th>Corn</th>
<th>Sugar Beets</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s = 1$ Above</td>
<td>170</td>
<td>80</td>
<td>250</td>
</tr>
<tr>
<td>$s = 2$ Average</td>
<td>425</td>
<td>240</td>
<td>5000</td>
</tr>
<tr>
<td>$s = 3$ Below</td>
<td>340</td>
<td>192</td>
<td>4000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Yield (T)</th>
<th>Sales (T)</th>
<th>Purchase (T)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s = 1$ Above</td>
<td>510</td>
<td>310</td>
</tr>
<tr>
<td></td>
<td>288</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>6000</td>
<td>6000</td>
</tr>
<tr>
<td>$s = 2$ Average</td>
<td>425</td>
<td>225</td>
</tr>
<tr>
<td></td>
<td>240</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>5000</td>
<td>5000</td>
</tr>
<tr>
<td>$s = 3$ Below</td>
<td>340</td>
<td>140</td>
</tr>
<tr>
<td></td>
<td>192</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td></td>
<td>48</td>
</tr>
</tbody>
</table>

Overall profit: $108,390
<table>
<thead>
<tr>
<th></th>
<th>Wheat</th>
<th>Corn</th>
<th>Sugar Beets</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>First Stage</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Area (acres)</td>
<td>170</td>
<td>80</td>
<td>250</td>
</tr>
<tr>
<td>s = 1 Above</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yield (T)</td>
<td>510</td>
<td>288</td>
<td>6000</td>
</tr>
<tr>
<td>Sales (T)</td>
<td>310</td>
<td>48</td>
<td>6000 (favor. price)</td>
</tr>
<tr>
<td>Purchase (T)</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>s = 2 Average</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yield (T)</td>
<td>425</td>
<td>240</td>
<td>5000</td>
</tr>
<tr>
<td>Sales (T)</td>
<td>225</td>
<td>–</td>
<td>5000 (favor. price)</td>
</tr>
<tr>
<td>Purchase (T)</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>s = 3 Below</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yield (T)</td>
<td>340</td>
<td>192</td>
<td>4000</td>
</tr>
<tr>
<td>Sales (T)</td>
<td>140</td>
<td>–</td>
<td>4000 (favor. price)</td>
</tr>
<tr>
<td>Purchase (T)</td>
<td>–</td>
<td>48</td>
<td>–</td>
</tr>
</tbody>
</table>

Overall profit: $108,390

**Observations:**
- Mario never sells beets at the unfavorable price;
- Sometimes he produces fewer beets than the quota;
- He always sells wheat, never having to buy it;
- He may sell or buy corn, depending on the year;
The value of perfect information

- Some of Mario’s decisions (e.g., underproducing beets, having to buy corn) would never take place if he had *perfect information*;
Some of Mario’s decisions (e.g., underproducing beets, having to buy corn) would never take place if he had perfect information; they appear because decisions have to be balanced against the various scenarios; his average long-term profit would be the average of the three profits, i.e., $115,456. The difference $115,456-$108,390=$7,016 is the expected value of perfect information (EVPI), i.e., the loss of profit due to the presence of uncertainty.
The value of perfect information

- Some of Mario’s decisions (e.g., underproducing beets, having to buy corn) would never take place if he had perfect information;
- They appear because decisions have to be balanced against the various scenarios;

Assume that the weather is cyclical (bad, fair, good) and Mario knows it.
- Then he would know what decisions to take every year (from the three original LPs).
The value of perfect information

- Some of Mario’s decisions (e.g., underproducing beets, having to buy corn) would never take place if he had *perfect information*;
- They appear because decisions have to be balanced against the various scenarios;

Assume that the weather is cyclical (bad, fair, good) and Mario knows it.
- Then he would know what decisions to take every year (from the three original LPs).
- His average long-term profit would be the average of the three profits, i.e., $115,456.

The difference $115,456-$108,390=$7,016 is the expected value of perfect information (EVPI), i.e., the loss of profit due to the presence of uncertainty.
The value of perfect information

Some of Mario’s decisions (e.g., underproducing beets, having to buy corn) would never take place if he had perfect information; they appear because decisions have to be balanced against the various scenarios.

Assume that the weather is cyclical (bad, fair, good) and Mario knows it.

Then he would know what decisions to take every year (from the three original LPs).

His average long-term profit would be the average of the three profits, i.e., $115,456.

The difference $115,456-$108,390=$7,016 is the expected value of perfect information (EVPI)

I.e., the loss of profit due to the presence of uncertainty.
The value of the stochastic solution

What would Mario’s profits in the long run be if he planted following the expected yields? (I.e., using the solution in the first original LP)
The value of the stochastic solution

What would Mario’s profits in the long run be if he planted following the expected yields? (I.e., using the solution in the first original LP)

• For each scenario, we can compute what the yields would be and what he would have to buy and/or could sell, and hence his profits in that year.
The value of the stochastic solution

What would Mario’s profits in the long run be if he planted following the expected yields? (I.e., using the solution in the first original LP)

- For each scenario, we can compute what the yields would be and what he would have to buy and/or could sell, and hence his profits in that year.
- The average profit is $107,240.
The value of the stochastic solution

What would Mario’s profits in the long run be if he planted following the expected yields? (I.e., using the solution in the first original LP)

- For each scenario, we can compute what the yields would be and what he would have to buy and/or could sell, and hence his profits in that year.
- The average profit is $107,240. The difference $108,390-107,240=1,150 is the value of the stochastic solution (VSS) I.e., the gain possible from solving the stochastic model
EVPI and VSS

- EVPI measures the value of knowing the future with certainty
- VSS measures the value of knowing and using distributions on future outcomes

In practice, EVPI is difficult to measure, so the emphasis is often on VSS.
Two-stage programs with fixed recourse

Mario’s problem is a *two-stage stochastic linear program with fixed recourse*.

**Definition**

Find

\[
\begin{align*}
\min z &= c^T x + \mathbb{E}[\min q(\omega)^T y(\omega)] \\
\text{s.t.} \quad Ax &= b \\
T(\omega)x + Wy(\omega) &= h(\omega) \\
x &\geq 0, \quad y(\omega) \geq 0
\end{align*}
\]

- \( \omega \) is a random event \( \omega \in \Omega \), which happens with probability \( p(\omega) \);
- \( x \) are the first-stage decisions, to be fixed *before* the realization of \( \omega \) is known;
Mario’s problem is a \textit{two-stage stochastic linear program with fixed recourse.}

**Definition**

Find

\[
\min z = c^T x + \mathbb{E}[\min q(\omega)^T y(\omega)]
\]

s.t. \(Ax = b\)

\[
T(\omega)x + Wy(\omega) = h(\omega)
\]

\(x \geq 0, \quad y(\omega) \geq 0\)

- \(\omega\) is a random event \(\omega \in \Omega\), which happens with probability \(p(\omega)\);
- \(x\) are the first-stage decisions, to be fixed \textit{before} the realization of \(\omega\) is known;
- Once we know the realization of \(\omega\), then we know the second-stage problem data \(q(\omega), h(\omega), \text{ and } T(\omega)\) (i.e., each component of these vector is a r.v.)
Two-stage programs with fixed recourse

Mario’s problem is a *two-stage stochastic linear program with fixed recourse*.

### Definition

Find

\[
\begin{align*}
\min z &= c^T x + \mathbb{E}[\min q(\omega)^T y(\omega)] \\
\text{s.t. } Ax &= b \\
T(\omega)x + Wy(\omega) &= h(\omega) \\
x \geq 0, & \quad y(\omega) \geq 0
\end{align*}
\]

- \(\omega\) is a random event \(\omega \in \Omega\), which happens with probability \(p(\omega)\);
- \(x\) are the first-stage decisions, to be fixed *before* the realization of \(\omega\) is known;
- Once we know the realization of \(\omega\), then we know the second-stage problem data \(q(\omega), h(\omega)\), and \(T(\omega)\) (i.e., each component of these vector is a r.v.)
- \(y(\omega)\) are the second stage decisions, which depend on \(\omega\) in the sense that they depend on the random constraints and costs;
Recourse

\[
\begin{align*}
\min z &= c^T x + \mathbb{E}[\min q(\omega)^T y(\omega)] \\
\text{s.t. } Ax &= b \\
T(\omega)x + Wy(\omega) &= h(\omega) \\
x &\geq 0, \quad y(\omega) \geq 0
\end{align*}
\]

- “Recourse” allows us to “correct” our first-stage decisions when additional information (i.e., \(\omega\)) is revealed.
Recourse

\[
\min z = c^T x + \mathbb{E}[\min q(\omega)^T y(\omega)]
\]

s.t. \( Ax = b \)

\[
T(\omega)x + W y(\omega) = h(\omega)
\]

\[
x \geq 0, \quad y(\omega) \geq 0
\]

- “Recourse” allows us to “correct” our first-stage decisions when additional information (i.e., \( \omega \)) is revealed.
- In fixed-recourse models, \( W \) is fixed, but in general it can be \( W(\omega) \)
Recourse

\[
\begin{align*}
\min z &= c^T x + \mathbb{E}[\min q(\omega)^T y(\omega)] \\
\text{s.t. } Ax &= b \\
T(\omega)x + Wy(\omega) &= h(\omega) \\
x &\geq 0, \quad y(\omega) \geq 0
\end{align*}
\]

- “Recourse” allows us to “correct” our first-stage decisions when additional information (i.e., \(\omega\)) is revealed.
- In fixed-recourse models, \(W\) is fixed, but in general it can be \(W(\omega)\).

E.g.,
- Think of an initial set of investments.
- As time passes we can then adjust our allocations to take into account changes in values.
- The cost/profits from the allocations depend on the random changes in values.
Second-stage problem

Given the first-stage decisions \( x \), we can define the second-stage or recourse problem:

\[
f(x, \omega) = \min Q(\omega)^T y(\omega)
\]

\[
Wy(\omega) = h(\omega) - T(\omega)x
\]

\[
y(\omega) \geq 0
\]
Second-stage problem

Given the first-stage decisions \( x \), we can define the second-stage or recourse problem:

\[
f(x, \omega) = \min Q(\omega)^T y(\omega)
\]

\[
Wy(\omega) = h(\omega) - T(\omega)x
\]

\[
y(\omega) \geq 0
\]

Let \( f(x) = \mathbb{E}[f(x, \omega)] \) be the second stage value function. Then we can write:

\[
\min c^T x + f(x)
\]

\[
Ax = b
\]

\[
x \geq 0
\]
Second-stage problem

Given the first-stage decisions \( x \), we can define the second-stage or recourse problem:

\[
\begin{align*}
f(x, \omega) &= \min Q(\omega)^T y(\omega) \\
Wy(\omega) &= h(\omega) - T(\omega)x \\
y(\omega) &\geq 0
\end{align*}
\]

Let \( f(x) = \mathbb{E}[f(x, \omega)] \) be the second stage value function. Then we can write:

\[
\begin{align*}
\min c^T x + f(x) \\
Ax &= b \\
x &\geq 0
\end{align*}
\]

- If we have complete information on \( f \), and \( f \) is linear, then the two-stage stochastic linear problem is just a deterministic linear program.
- The crux of the matter is in \( f(x) \).
Scenarios

- Assume $\Omega = \{\omega_1, \ldots, \omega_S\}$ (scenarios) and let $p = (p_1, \ldots, p_s)$ be the probability distribution on $\Omega$;
Scenarios

• Assume $\Omega = \{\omega_1, \ldots, \omega_S\}$ (scenarios) and let $p = (p_1, \ldots, p_s)$ be the probability distribution on $\Omega$;

• Then

$$\mathbb{E}[\min q(\omega)^T y(\omega)] = \sum_{k=1}^{S} p_k \min_{y(\omega_k)} q(\omega_k)^T y(\omega_k)$$
Scenarios

- Assume $\Omega = \{\omega_1, \ldots, \omega_S\}$ (scenarios) and let $p = (p_1, \ldots, p_s)$ be the probability distribution on $\Omega$;
- Then

$$\mathbb{E}[\min q(\omega)^T y(\omega)] = \sum_{k=1}^{S} p_k \min_{y(\omega_k)} q(\omega_k)^T y(\omega_k)$$

- And we can rewrite the stochastic program as:

$$\min z = c^T x + \sum_{k=1}^{S} p_k \min_{y_k} q_k^T y_k$$

s.t. $Ax = b$

$$T_k x + W_k y_k = h_k \text{ for } k = 1, \ldots, S$$

$x \geq 0$

$y_k \geq 0$ for $k = 1 \ldots, S$

- We now have a different second stage decision vector $y_k$ for each scenario $k$;
The deterministic equivalent problem

\[
\begin{align*}
\min_{x, y_1, \ldots, y_k} & \quad c^T x + p_1 q_1^T y_1 + \cdots + p_s q_s^T y_s \\
\text{s.t.} & \quad Ax = b \\
& \quad B_1 x + W_1 y_i = h_1 \\
& \quad \vdots = \vdots \\
& \quad B_s x + W_s y_s = h_s \\
& \quad x, y_1, \ldots, y_s \geq 0
\end{align*}
\]
The deterministic equivalent problem

\[
\min_{x, y_i, \ldots, y_k} \left( c^T x + p_1 q_1^T y_1 + \cdots + p_S q_S^T y_S \right)
\]

\[
A x = b
\]
\[
B_1 x + W_1 y_i = h_1
\]
\[
\vdots\]
\[
B_S x + W_S y_S = h_S
\]
\[
x, y_1, \ldots, y_s \geq 0
\]

- There are \( S \) copies of the second stage decision variables: it may be very big.
The deterministic equivalent problem

\[
\min_{x, y_1, \ldots, y_k} c^T x + p_1 q_1^T y_1 + \cdots + ps q_s^T y_s
\]
\[
Ax + W_1 y_i = h_1
\]
\[
B_1 x + W_1 y_i = h_1
\]
\[
\vdots
\]
\[
B_s x + W_s y_s = h_s
\]
\[
x, y_1, \ldots, y_s \geq 0
\]

• There are \( S \) copies of the second stage decision variables: it may be very big.
• The problem has a very nice structure. We should be able to exploit it.
Multi-stage programs with recourse

- The recourse decisions can be made at $n \geq 2$ points in time, called stages;
Multi-stage programs with recourse

- The recourse decisions can be made at $n \geq 2$ points in time, called stages;
- The random event $\omega$ is a vector $(o_1, \ldots, o_{n-1})$ that gets revealed progressively over time;
Multi-stage programs with recourse

- The recourse decisions can be made at \( n \geq 2 \) points in time, called *stages*;
- The random event \( \omega \) is a vector \((o_1, \ldots, o_{n-1})\) that gets revealed progressively over time;
- First stage decisions are taken before any component of \( \omega \) is revealed;
Multi-stage programs with recourse

- The recourse decisions can be made at \( n \geq 2 \) points in time, called stages;
- The random event \( \omega \) is a vector \((o_1, \ldots, o_{n-1})\) that gets revealed progressively over time;
- First stage decisions are taken before any component of \( \omega \) is revealed;
- Then \( o_1 \) is revealed and the second stage decisions are taken...
Multi-stage programs with recourse

• The recourse decisions can be made at $n \geq 2$ points in time, called stages;
• The random event $\omega$ is a vector $(o_1, \ldots, o_{n-1})$ that gets revealed progressively over time;
• First stage decisions are taken before any component of $\omega$ is revealed;
• Then $o_1$ is revealed and the second stage decisions are taken
• Then $o_2$ and so on, alternating between revealing a new component and taking the current stage decisions.
Scenarios tree

- Assume $\Omega = \{\omega_1, \ldots, \omega_S\}$
- Some scenarios may be identical in their first components;
- They “become” differentiated in later stages;

![Scenario tree](image-url)
Scenarios tree

- Assume $\Omega = \{\omega_1, \ldots, \omega_5\}$
- Some scenarios may be identical in their first components;
- They “become” differentiated in later stages;
- We can represent this situation as a scenario tree

![Scenario Tree Diagram]

Stage

1
2
3

4 scenarios
Properties of the scenario tree

- Nodes are labeled from 1 to $N$, with 1 being the root.
- Each node $i$ in stage $k \geq 2$ has a single mother $m(i)$.
- The paths from the root to the leaves represent scenarios.
- The scenarios that pass through node $i$ in stage $k$ have identical components $o_1$ to $o_{k-1}$.
Formulation of a multi-stage stochastic problem

- For each node $i$ of the tree, there is a recourse decision vector $x_i$;
For each node $i$ of the tree, there is a recourse decision vector $x_i$;
• For each node $i$, let $r_i$ be the sum of the $p_k$ for the scenarios $\omega_k$ that go through $i$;
Formulation of a multi-stage stochastic problem

- For each node $i$ of the tree, there is a recourse decision vector $x_i$;
- For each node $i$, let $r_i$ be the sum of the $p_k$ for the scenarios $\omega_k$ that go through $i$;

Multi-stage stochastic program with recourse:

$$\min_{x_1, \ldots, x_n} \sum_{i=1}^{n} r_i c_i^T x_i$$

$$Ax_1 = b$$

$$B_i x_{a(i)} + W_i x_i = h_i \text{ for } i = 2, \ldots, N$$

$$x_i \geq 0$$
Example

- $r_4 = p_1$, $r_5 = p_2$, $r_6 = p_3$, $r_7 = p_4$
- $r_2 = p_1 + p_2 + p_3$, $r_3 = p_4$, $r_2 + r_3 = 1 = r_1$. 
Example

- \( r_4 = p_1, r_5 = p_2, r_6 = p_3, r_7 = p_4 \)
- \( r_2 = p_1 + p_2 + p_3, r_3 = p_4, r_2 + r_3 = 1 = r_1. \)

\[
\begin{align*}
\min & \quad c^T x_1 + r_2 q_2^T x_2 + \cdots + r_7 q_7^T x_7 \\
A x_1 & = b \\
T_2 x_1 + W_2 x_2 & = h_2, T_3 x_1 + W_3 x_3 = h_3 \\
T_4 x_2 + W_4 x_4 & = h_4, T_5 x_2 + W_5 x_5 = h_5, T_6 x_2 + W_6 x_6 = h_6 \\
T_7 x_3 + W_7 x_7 & = h_7 \\
x_i & \geq 0
\end{align*}
\]

4 scenarios
Observations

\[
\begin{align*}
\min \ & c^T x_1 + r_2 q_2^T x_2 + \cdots + r_7 q_7^T x_7 \\
A x_1 & = b \\
T_2 x_1 + W_2 x_2 & = h_2, \ T_3 x_1 + W_3 x_3 = h_3 \\
T_4 x_2 + W_4 x_4 & = h_4, \ T_5 x_2 + W_5 x_5 = h_5, \ T_6 x_2 + W_6 x_6 = h_6 \\
T_7 x_3 + W_7 x_7 & = h_7 \\
x_i & \geq 0
\end{align*}
\]

- The size of the LP increases rapidly with the number of stages (e.g., 10 stages and binary tree means 1024 scenarios, 2047 decision vectors, 2048 constraints);
Observations

\[
\begin{align*}
\min & \quad c^T x_1 + r_2 q_2^T x_2 + \cdots + r_7 q_7^T x_7 \\
A x_1 & = b \\
T_2 x_1 + W_2 x_2 & = h_2, \quad T_3 x_1 + W_3 x_3 = h_3 \\
T_4 x_2 + W_4 x_4 & = h_4, \quad T_5 x_2 + W_5 x_5 = h_5, \quad T_6 x_2 + W_6 x_6 = h_6 \\
T_7 x_3 + W_7 x_7 & = h_7 \\
x_i & \geq 0
\end{align*}
\]

- The size of the LP increases rapidly with the number of stages (e.g., 10 stages and binary tree means 1024 scenarios, 2047 decision vectors, 2048 constraints);
- The problem still has the “nice” structure of 2-stage stochastic problems;
Example from personal finance

- We have $55k to invest today in bonds or stocks;
Example from personal finance

- We have $55k to invest today in bonds or stocks;
- 15 years from now, we would like to have at least \( G = $80k \) to pay for a 2-year master program;
- If we don’t have $80k after 15 years, we can borrow for a cost of 4% the amount we need;
Example from personal finance

- We have $55k to invest today in bonds or stocks;
- 15 years from now, we would like to have at least $G=$80k to pay for a 2-year master program;
- If we don’t have $80k after 15 years, we can borrow for a cost of 4% the amount we need;
- We can change the investments every 5 years, so we have 3 investment periods.
Example from personal finance

• We have $55k to invest today in bonds or stocks;
• 15 years from now, we would like to have at least $G=80k$ to pay for a 2-year master program;
• If we don’t have $80k after 15 years, we can borrow for a cost of 4% the amount we need;
• We can change the investments every 5 years, so we have 3 investment periods.
• We assume that over the 3 decision periods, 8 scenarios are possible and all equally likely ($p_i = 0.125$):
  over each 5-year period, either stocks give a return of 1.25 and bonds of 1.14, or stocks give a return of 1.06 and bonds a return of 1.12
• We would like to maximize the expected amount of money we have left at the end of the 9 years (taking into account the eventual costs of borrowing and our tuition expenses)
22 1 Introduction and Examples

$s = 1, 2, 5, 6,$ and for $t = 3, s = 1, 3, 5, 7$. In the other cases, $\xi(1, t, s) = 1.06$, $\xi(2, t, s) = 1.12$.

Fig. 3 Tree of scenarios for three periods. The eight scenarios are represented by the tree in Figure 3. The scenario tree divides into branches corresponding to different realizations of the random returns. Because Scenarios 1 to 4, for example, have the same return for $t = 1$, they all follow the same first branch. Scenarios 1 and 2 then have the same second branch and finally divide completely in the last period. To show this more explicitly, we may refer to each scenario by the history of returns indexed by $s$ for periods $t = 1, 2, 3$ as indicated on the tree in Figure 3. In this way, Scenario 1 may also be represented as $(s_1, s_2, s_3) = (1, 1, 1)$. With the tree representation, we need only have a decision vector for each node of the tree. The decisions at $t = 1$ are $x(1, 1)$ and $x(2, 1)$ for the amounts invested in stocks (1) and bonds (2) at the outset. For $t = 2$, we would have $x(i, 2, s_1)$ where $i = 1, 2$ for the type of investment and $s_1 = 1, 2$ for the first-period return outcome. Similarly, the decisions at $t = 3$ are $x(i, 3, s_1, s_2)$.

With these decision variables defined, we can formulate a mathematical program to maximize expected utility. Because the concave utility function in Figure 1 is piecewise linear, we just need to define deficit or shortage and excess or surplus variables, $w(i_1, i_2, i_3)$ and $y(i_1, i_2, i_3)$, and we can maintain an infeasible model. The objective is simply a probability- and penalty-weighted sum of these terms, which, in general, becomes:
1.2 Financial Planning and Control

The first-period constraint is simply to invest the initial wealth:

\[ x(i, t, s_1, \ldots, s_{t-1}) \geq 0, \quad y(s_1, s_2, s_3) \geq 0, \quad w(s_1, s_2, s_3) \geq 0, \]

for all \( i, t, s_1, s_2, s_3 \).

\[ \text{max } z = \sum_{s_1 = 1}^{2} \sum_{s_2 = 1}^{2} \sum_{s_3 = 1}^{2} 0.125(y(s_1, s_2, s_3) - 4w(s_1, s_2, s_3)) \]

s. t.

\[ x(1, 1) + x(2, 1) = 55, \]

\[ -1.25x(1, 1) - 1.14x(2, 1) + x(1, 2, 1) + x(2, 2, 1) = 0, \]

\[ -1.06x(1, 1) - 1.12x(2, 1) + x(1, 2, 2) + x(2, 2, 2) = 0, \]

\[ -1.25x(1, 2, 1) - 1.14x(2, 2, 1) + x(1, 3, 1, 1) + x(2, 3, 1, 1) = 0, \]

\[ -1.06x(1, 2, 1) - 1.12x(2, 2, 1) + x(1, 3, 1, 2) + x(2, 3, 1, 2) = 0, \]

\[ -1.25x(1, 2, 2) - 1.14x(2, 2, 2) + x(1, 3, 2, 1) + x(2, 3, 2, 1) = 0, \]

\[ -1.06x(1, 2, 2) - 1.12x(2, 2, 2) + x(1, 3, 2, 2) + x(2, 3, 2, 2) = 0, \]

\[ 1.25x(1, 3, 1, 1) + 1.14x(2, 3, 1, 1) - y(1, 1, 1) + w(1, 1, 1) = 80, \]

\[ 1.06x(1, 3, 1, 1) + 1.12x(2, 3, 1, 1) - y(1, 1, 2) + w(1, 1, 2) = 80, \]

\[ 1.25x(1, 3, 1, 2) + 1.14x(2, 3, 1, 2) - y(1, 2, 1) + w(1, 2, 1) = 80, \]

\[ 1.06x(1, 3, 1, 2) + 1.12x(2, 3, 1, 2) - y(1, 2, 2) + w(1, 2, 2) = 80, \]

\[ 1.25x(1, 3, 2, 1) + 1.14x(2, 3, 2, 1) - y(2, 1, 1) + w(2, 1, 1) = 80, \]

\[ 1.06x(1, 3, 2, 1) + 1.12x(2, 3, 2, 1) - y(2, 1, 2) + w(2, 1, 2) = 80, \]

\[ 1.25x(1, 3, 2, 2) + 1.14x(2, 3, 2, 2) - y(2, 2, 1) + w(2, 2, 1) = 80, \]

\[ 1.06x(1, 3, 2, 2) + 1.12x(2, 3, 2, 2) - y(2, 2, 2) + w(2, 2, 2) = 80, \]
Optimal solution

<table>
<thead>
<tr>
<th>Period, Scenario</th>
<th>Stock</th>
<th>Bonds</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,1-8</td>
<td>41.5</td>
<td>13.5</td>
</tr>
<tr>
<td>2,1-4</td>
<td>65.1</td>
<td>2.17</td>
</tr>
<tr>
<td>2,5-8</td>
<td>36.7</td>
<td>22.4</td>
</tr>
<tr>
<td>3,1-2</td>
<td>83.8</td>
<td>0.00</td>
</tr>
<tr>
<td>3,3-4</td>
<td>0.00</td>
<td>71.4</td>
</tr>
<tr>
<td>3,5-6</td>
<td>0.00</td>
<td>71.4</td>
</tr>
<tr>
<td>3,7-8</td>
<td>64.0</td>
<td>0.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Above G</th>
<th>Below G</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>24.8</td>
<td>0.00</td>
</tr>
<tr>
<td>2</td>
<td>8.87</td>
<td>0.00</td>
</tr>
<tr>
<td>3</td>
<td>1.43</td>
<td>0.00</td>
</tr>
<tr>
<td>4</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>5</td>
<td>1.43</td>
<td>0.00</td>
</tr>
<tr>
<td>6</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>7</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>8</td>
<td>0.00</td>
<td>12.2</td>
</tr>
</tbody>
</table>

- The initial solution is heavy in stocks;
**Optimal solution**

<table>
<thead>
<tr>
<th>Period, Scenario</th>
<th>Stock</th>
<th>Bonds</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,1-8</td>
<td>41.5</td>
<td>13.5</td>
</tr>
<tr>
<td>2,1-4</td>
<td>65.1</td>
<td>2.17</td>
</tr>
<tr>
<td>2,5-8</td>
<td>36.7</td>
<td>22.4</td>
</tr>
<tr>
<td>3,1-2</td>
<td>83.8</td>
<td>0.00</td>
</tr>
<tr>
<td>3,3-4</td>
<td>0.00</td>
<td>71.4</td>
</tr>
<tr>
<td>3,5-6</td>
<td>0.00</td>
<td>71.4</td>
</tr>
<tr>
<td>3,7-8</td>
<td>64.0</td>
<td>0.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Above G</th>
<th>Below G</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>24.8</td>
<td>0.00</td>
</tr>
<tr>
<td>2</td>
<td>8.87</td>
<td>0.00</td>
</tr>
<tr>
<td>3</td>
<td>1.43</td>
<td>0.00</td>
</tr>
<tr>
<td>4</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>5</td>
<td>1.43</td>
<td>0.00</td>
</tr>
<tr>
<td>6</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>7</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>8</td>
<td>0.00</td>
<td>12.2</td>
</tr>
</tbody>
</table>

- The initial solution is heavy in stocks;
- After the first period, we become either even more unbalanced towards stocks, or try to rebalance;
Optimal solution

<table>
<thead>
<tr>
<th>Period, Scenario</th>
<th>Stock</th>
<th>Bonds</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,1-8</td>
<td>41.5</td>
<td>13.5</td>
</tr>
<tr>
<td>2,1-4</td>
<td>65.1</td>
<td>2.17</td>
</tr>
<tr>
<td>2,5-8</td>
<td>36.7</td>
<td>22.4</td>
</tr>
<tr>
<td>3,1-2</td>
<td>83.8</td>
<td>0.00</td>
</tr>
<tr>
<td>3,3-4</td>
<td>0.00</td>
<td>71.4</td>
</tr>
<tr>
<td>3,5-6</td>
<td>0.00</td>
<td>71.4</td>
</tr>
<tr>
<td>3,7-8</td>
<td>64.0</td>
<td>0.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Above G</th>
<th>Below G</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>24.8</td>
<td>0.00</td>
</tr>
<tr>
<td>2</td>
<td>8.87</td>
<td>0.00</td>
</tr>
<tr>
<td>3</td>
<td>1.43</td>
<td>0.00</td>
</tr>
<tr>
<td>4</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>5</td>
<td>1.43</td>
<td>0.00</td>
</tr>
<tr>
<td>6</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>7</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>8</td>
<td>0.00</td>
<td>12.2</td>
</tr>
</tbody>
</table>

- The initial solution is heavy in stocks;
- After the first period, we become either even more unbalanced towards stocks, or try to rebalance;
- The final investments are either all in stocks or all in bonds;
The initial solution is heavy in stocks;
After the first period, we become either even more unbalanced towards stocks, or try to rebalance;
The final investments are either all in stocks or all in bonds;
Despite having to borrow money only once, because of the cost associated to it, the expected utility is negative: -$1,514.
Value of the stochastic solution

- What if we used a deterministic model replacing the random returns with their expectation?

\[ V_{SS} = -1,514 - (-3,788) = 2,274 \]
Value of the stochastic solution

- What if we used a deterministic model replacing the random returns with their expectation?
- The expected return of stock is $1.155$ in each period, while bonds return only $1.113$
Value of the stochastic solution

- What if we used a deterministic model replacing the random returns with their expectation?
- The expected return of stock is 1.155 in each period, while bonds return only 1.113
- The optimal investment plan would place all funds in stocks in each period
Value of the stochastic solution

- What if we used a deterministic model replacing the random returns with their expectation?
- The expected return of stock is $1.155$ in each period, while bonds return only $1.113$
- The optimal investment plan would place all funds in stocks in each period
- The resulting utility is -$3,788$. 

\[VSS = -1.514 - (-3.788) = 2.274\]
Value of the stochastic solution

- What if we used a deterministic model replacing the random returns with their expectation?
- The expected return of stock is $1.155$ in each period, while bonds return only $1.113$
- The optimal investment plan would place all funds in stocks in each period
- The resulting utility is $-3,788$.

\[ VSS = -1,514 - (-3,788) = 2,274 \]
Solving stochastic problems

Let’s consider a two-stage problem:

\[
\begin{align*}
\min_{x, y_i, \ldots, y_k} & \quad c^T x + p_1 q_1^T y_1 + \cdots + p_S q_S^T y_S \\
Ax & \quad = b \\
B_1 x + W_1 y_i & \quad = h_1 \\
\vdots & \quad = \vdots \\
B_S x & \quad + W_S y_S = h_S \\
x, y_1, \ldots & \quad , y_S \geq 0
\end{align*}
\]
Benders decomposition

- The size of the deterministic equivalent problem depends on:
  - the number of decision stages; and
  - the branching factor at each node.
Benders decomposition

- The size of the deterministic equivalent problem depends on
  - the number of decision stages; and
  - the branching factor at each node
- It is a limiting factor in solving realistic problems
- We must take advantage of the special structure of the LP
Benders decomposition

- The size of the deterministic equivalent problem depends on
  - the number of decision stages; and
  - the branching factor at each node
- It is a limiting factor in solving realistic problems
- We must take advantage of the special structure of the LP
- Benders decomposition, (aka the L-Shaped method) solves a
  number of smaller LPs, leveraging the structure.
Intuition

- Solve a “master” problem involving only $x$ and the constraint $Ax = b$;
- Solve a series of independent “recourse problems”, each involving a different vector of variables $y_k$.
Intuition

• Solve a “master” problem involving only $x$ and the constraint $Ax = b$;
• Solve a series of independent “recourse problems”, each involving a different vector of variables $y_k$
• All the problems are linear
• The problems are much smaller
Intuition

- Solve a “master” problem involving only $x$ and the constraint $Ax = b$;
- Solve a series of independent “recourse problems”, each involving a different vector of variables $y_k$;
- All the problems are linear;
- The problems are much smaller;
- The recourse problems are solved for a given vector $x$;
- Their solutions are used to generate inequalities that are added to the master problem.
Intuition

• Solve a “master” problem involving only $x$ and the constraint $Ax = b$;
• Solve a series of independent “recourse problems”, each involving a different vector of variables $y_k$
• All the problems are linear
• The problems are much smaller
• The recourse problems are solved for a given vector $x$
• Their solutions are used to generate inequalities that are added to the master problem
• This create a new master problem, which we solve to obtain a new $x$, and the method iterates;
We can rewrite the two-stage problem as:

$$\max_x c^T x + P_1(x) + \cdots + P_S(x)$$

$$Ax = b$$

$$x \geq 0$$

where, for $k = 1, \ldots, S$:

$$P_k(x) = \max_{y_k} p_k q_k^T y_k$$

$$W_k y_k = h_k - B_k x$$

$$y_k \geq 0$$
The recourse linear problem

\[ P_k(x) = \max_{y_k} p_k q_k^T y_k \]
\[ W_k y_k = h_k - B_k x \]
\[ y_k \geq 0 \]

- The recourse linear problems \( P_k(x), k = 1, \ldots, S \) will be solved for a sequence of vectors \( x^i, i = 0, \ldots \);
The recourse linear problem

\[ P_k(x) = \max_{y_k} p_k q_k^T y_k \]

\[ W_k y_k = h_k - B_k x \]

\[ y_k \geq 0 \]

- The recourse linear problems \( P_k(x), \ k = 1, \ldots, S \) will be solved for a sequence of vectors \( x^i, \ i = 0, \ldots; \)
- The initial vector \( x^0 \) can be obtained by solving

\[ \max_x c^T x \]

\[ Ax = b \]

\[ x \geq 0 \]
The recourse linear problem

\[ P_k(x) = \max_{y_k} p_k q_k^T y_k \]

\[ W_k y_k = h_k - B_k x \]

\[ y_k \geq 0 \]

- The recourse linear problems \( P_k(x), k = 1, \ldots, S \) will be solved for a sequence of vectors \( x^i, i = 0, \ldots \);
- The initial vector \( x^0 \) can be obtained by solving

\[ \max_x c^T x \]

\[ Ax = b \]

\[ x \geq 0 \]

Observations:

- \( x^0 \) may not be optimal for the original problem.
- \( x^0 \) may make some of the recourse problems infeasible.
The recourse linear problem

\[ P_k(x) = \max_{y_k} p_k q_k^T y_k \]
\[ W_k y_k = h_k - B_k x \]
\[ y_k \geq 0 \]

- The recourse linear problems \( P_k(x), k = 1, \ldots, S \) will be solved for a sequence of vectors \( x^i, i = 0, \ldots \); 
- The initial vector \( x^0 \) can be obtained by solving

\[ \max_x c^T x \]
\[ Ax = b \]
\[ x \geq 0 \]

Observations:
- \( x^0 \) may not be optimal for the original problem.
The recourse linear problem

\[ P_k(x) = \max_{y_k} p_k q_k^T y_k \]
\[ W_k y_k = h_k - B_k x \]
\[ y_k \geq 0 \]

- The recourse linear problems \( P_k(x), \ k = 1, \ldots, S \) will be solved for a sequence of vectors \( x^i, \ i = 0, \ldots \);
- The initial vector \( x^0 \) can be obtained by solving

\[ \max_x c^T x \]
\[ A x = b \]
\[ x \geq 0 \]

Observations:
- \( x^0 \) may not be optimal for the original problem.
- \( x^0 \) may make some of the recourse problems infeasible.
Suppose that we have computed $x^i$ from the master program; The dual of a recourse problem $P_k(x)$, given $x^i$ is:

$$P_k(x^i) = \min_{u_k} u_k^T (h_k - T_k x^i)$$

$$W_k^T u_k \geq p_k q_k$$
Suppose that we have computed $x^i$ from the master program; The dual of a recourse problem $P_k(x)$, given $x^i$ is:

$$P_k(x^i) = \min_{u_k} u_k^T (h_k - T_k x^i)$$

$$W_k^T u_k \geq p_k q_k$$

Assume that the primal recourse has optimal solution $y^i_k$ and that $u^k_i$ is the corresponding optimal dual. Then we have

$$P_k(x^i) = (u^i_k)^T (h_k - T_k x^i)$$
Suppose that we have computed \( x^i \) from the master program; The dual of a recourse problem \( P_k(x) \), given \( x^i \) is:

\[
P_k(x^i) = \min_{u_k^i} u_k^T (h_k - T_k x^i)
\]

\[
W_k^T u_k \geq p_k q_k
\]

Assume that the primal recourse has optimal solution \( y^i_k \) and that \( u_k^i \) is the corresponding optimal dual. Then we have

\[
P_k(x^i) = (u_k^i)^T (h_k - T_k x^i)
\]

From the fact that \( x^i \) is not necessarily optimal for the original problem, we have that, for an optimal solution \( x^* \) of the original problem,

\[
P_k(x^*) \leq (u_k^i)^T (h_k - T_k x^*)
\]
Suppose that we have computed $x^i$ from the master program; The dual of a recourse problem $P_k(x)$, given $x^i$ is:

$$P_k(x^i) = \min_{u_k} u_k^T (h_k - T_k x^i)$$

$$W_k^T u_k \geq p_k q_k$$

Assume that the primal recourse has optimal solution $y^i_k$ and that $u^i_k$ is the corresponding optimal dual. Then we have

$$P_k(x^i) = (u_k^i)^T (h_k - T_k x^i)$$

From the fact that $x^i$ is not necessarily optimal for the original problem, we have that, for an optimal solution $x^*$ of the original problem,

$$P_k(x^*) \leq (u_k^i)^T (h_k - T_k x^*)$$

Subtract the 1st expression from the 2nd. We can then add the following optimality cut to the current master linear program:

$$P_k(x) \leq (u_k^i)^T (T_k x^i - T_k x) + P_k(x^i)$$
• If the primal recourse is infeasible, then the dual is unbounded
• If the primal recourse is infeasible, then the dual is unbounded

• We are only interested in first-stage decisions $x$ that leads to *feasible* second stage decisions $y_k$;
• If the primal recourse is infeasible, then the dual is unbounded
• We are only interested in first-stage decisions $x$ that leads to feasible second stage decisions $y_k$;
• Let then $u_k^i$ be a direction where the dual is unbounded, i.e.:

$$(u_k^i)^T(h_k - T_k x^i) \leq 0 \quad \text{and} \quad W_k^T u_k^i \geq p_k q_k$$
• If the primal recourse is infeasible, then the dual is unbounded
• We are only interested in first-stage decisions $x$ that leads to feasible second stage decisions $y_k$;
• Let then $u^i_k$ be a direction where the dual is unbounded, i.e.:

$$ (u^i_k)^T(h - T_k x^i) \leq 0 \quad \text{and} \quad W^T_k u^i_k \geq p_k q_k $$

• We can add the following feasibility cut to the current master program:

$$ (u^i_k)^T(h - T_k x) \geq 0 $$
After solving the recourse problems for each \( k \) we have a lower bound to the optimal value of the stochastic program:

\[
LB = c^T x^i + P_1(x^i) + \cdots + P_S(x^i)
\]

where \( P_k(x^i) = -\infty \) if the corresponding problem is infeasible.
After adding all the optimality and feasibility cuts found so far (for \( j = 0, \ldots, i \)) to the master program, we obtain a new linear program:

\[
\begin{align*}
\max_{x, z_1, \ldots, z_S} & \quad c^T x + \sum_{k=1}^{S} z_k \\
A x & = b \\
z_k & \leq (u^j_k)^T (T_k x^j - T_k x) + P_k(x^j) \text{ for some pairs } (j, k) \\
0 & \leq (u^j_k)^T (h_k - T_k x) \text{ for the remaining pairs} \\
x & \geq 0
\end{align*}
\]
After adding all the optimality and feasibility cuts found so far (for \( j = 0, \ldots, i \)) to the master program, we obtain a new linear program:

\[
\begin{align*}
\max_{x, z_1, \ldots, z_S} & \quad c^T x + \sum_{k=1}^{S} z_k \\
Ax &= b \\
z_k &\leq (u^j_k)^T(T_k x^j - T_k x) + P_k(x^j) \text{ for some pairs } (j, k) \\
0 &\leq (u^j_k)^T(h_k - T_k x) \text{ for the remaining pairs} \\
x &\geq 0
\end{align*}
\]

By solving this problem we obtain:

- new first decision variables \( x^{i+1} \); and
- a new upper bound \( UB \) to the optimal value of the original stochastic problem
After adding all the optimality and feasibility cuts found so far (for \( j = 0, \ldots, i \)) to the master program, we obtain a new linear program:

\[
\max_{x, z_1, \ldots, z_S} \quad c^T x + \sum_{k=1}^{S} z_k \\
Ax = b \\
z_k \leq (u_k^j)^T (T_k x^j - T_k x) + P_k(x^j) \text{ for some pairs } (j, k) \\
0 \leq (u_k^j)^T (h_k - T_k x) \text{ for the remaining pairs} \\
x \geq 0
\]

By solving this problem we obtain:

- new first decision variables \( x^{i+1} \); and
- a new upper bound \( UB \) to the optimal value of the original stochastic problem

Bender decomposition stops when \( LB \) and \( UB \) are closer than a desired threshold.
Algorithm

• Start with the “simplest” master program to obtain $x_0$.
• Set $UB = +\infty$ and $LB = -\infty$.
• While $UB - LB > \text{threshold}$:
  • For $k = 1, \ldots, S$
    • Try to solve recourse problem $P_k$.
    • If feasible, add optimality cut
    • Otherwise, add feasibility cut
  • Compute new $LB$ using $x^i$
  • Solve the new master program and obtain the new $UB$
Benders decomposition can be used for multi-stage problems:

- the stages are partitioned into:
  - a first set corresponding to the master problem; and
  - a second set corresponding to the recourse problems
- When the first set variables are fixed, then we have separate linear problems for each stage in the remaining set
- Solving these LPs gives additional feasibility cuts and optimality cuts

Benders composition is very easily parallelizable, thanks to the independence of the recourse problems
Theory versus practice

- We assumed to know the probability distribution over the random parameters.
Theory versus practice

• We assumed to know the probability distribution over the random parameters.

• In multistage programs, we also assumed a finite number of scenarios.

We now discuss how to “work around” these assumptions in practice.
Modeling the random parameters

\[
\begin{align*}
\min z &= c^T x + \mathbb{E}[\min q(\omega)^T y(\omega)] \\
\text{s.t. } Ax &= b \\
T(\omega)x + Wy(\omega) &= h(\omega) \\
x &\geq 0, \quad y(\omega) \geq 0
\end{align*}
\]

The question is:

How can we model the random parameters \( q(\omega) \) and \( T(\omega) \)?

I.e., how do we study their correlation and evolution in time.
Modeling the random parameters

\[
\begin{align*}
\min z &= c^T x + \mathbb{E}[\min q(\omega)^T y(\omega)] \\
\text{s.t. } Ax &= b \\
T(\omega)x + Wy(\omega) &= h(\omega) \\
x &\geq 0, \quad y(\omega) \geq 0
\end{align*}
\]

The question is:

How can we model the random parameters \( q(\omega) \) and \( T(\omega) \)?

I.e., how do we study their correlation and evolution in time.

• We want a generative model that can be fitted using historical data
Modeling the random parameters

\[
\min z = c^T x + \mathbb{E}[\min q(\omega)^T y(\omega)]
\]
\[
\text{s.t. } Ax = b
\]
\[
T(\omega)x + W y(\omega) = h(\omega)
\]
\[
x \geq 0, \quad y(\omega) \geq 0
\]

The question is:

How can we model the random parameters \( q(\omega) \) and \( T(\omega) \)?

I.e., how do we study their correlation and evolution in time.

- We want a generative model that can be fitted using historical data
- Later, we sample from the generative model to create time series that we use to create the scenario tree.
A possible approach is to fit a autoregressive model using historical data.
A possible approach is to fit an autoregressive model using historical data. Let $r_t$ be the random vector of parameters at time $t$. 

A possible approach is to fit a *autoregressive* model using historical data. Let $r_t$ be the random vector of parameters at time $t$. 

The autoregressive model is defined as:

$$
r_t = D_0 + D_1 r_{t-1} + \cdots + D_p r_{t-p} + \epsilon_t$$

- $p$ is the number of lags modeling the autoregression.
- $D_0, D_1, \ldots, D_p$ are the time-independent constant matrices. They are estimated from historical data using, e.g., maximum likelihood, least squares, etc.
- $\epsilon_t$ is a vector of i.i.d. random perturbations with zero mean.
A possible approach is to fit a autoregressive model using historical data. Let $r_t$ be the random vector of parameters at time $t$. We model $r_t$ as:

$$r_t = D_0 + D_1 r_{t-1} + \cdots + D_p r_{t-p} + \varepsilon_t$$
A possible approach is to fit an autoregressive model using historical data. Let \( r_t \) be the random vector of parameters at time \( t \). We model \( r_t \) as:

\[
    r_t = D_0 + D_1 r_{t-1} + \cdots + D_p r_{t-p} + \varepsilon_t
\]

- \( p \) is the number of lags modeling the autoregression.
Autoregressive model

A possible approach is to fit an autoregressive model using historical data. Let $r_t$ be the random vector of parameters at time $t$. We model $r_t$ as:

$$r_t = D_0 + D_1 r_{t-1} + \cdots + D_p r_{t-p} + \varepsilon_t$$

- $p$ is the number of lags modeling the autoregression.
- $D_0, D_1, \ldots, D_p$ are the time-independent constant matrices. They are estimated from historical data using, e.g., maximum likelihood, least squares, etc.
- $\varepsilon$ is a vector of i.i.d. random perturbations with zero mean.
Constructing scenario trees

• We assume to know the exact probability distribution;
Constructing scenario trees

- We assume to know the exact probability distribution;
- The scenario approach inherently implies a discrete probability distribution;
Constructing scenario trees

- We assume to know the exact probability distribution;
- The scenario approach inherently implies a discrete probability distribution;
- The number of scenarios increases very fast with the number of stages and the branching factor of the scenario tree, leading to computationally expensive formulations;
- Too few scenarios may lead to very large errors in our results.
Constructing scenario trees

- We assume to know the exact probability distribution;
- The scenario approach inherently implies a discrete probability distribution;
- The number of scenarios increases very fast with the number of stages and the branching factor of the scenario tree, leading to computationally expensive formulations;
- Too few scenarios may lead to very large errors in our results.

We must learn the probability distribution from existing data and somehow fit it into a tree with low branching factor.
Random sampling

• We can generate scenarios directly from the autoregressive model:

$$r_t = D_0 + D_1 r_{t-1} + \cdots + D_p r_{t-p} + \varepsilon_t$$

where $\varepsilon_t \sim N(0, \Sigma)$ is a vector of random errors.
Random sampling

• We can generate scenarios directly from the autoregressive model:

\[ r_t = D_0 + D_1 r_{t-1} + \cdots + D_p r_{t-p} + \varepsilon_t \]

where \( \varepsilon_t \sim N(0, \Sigma) \) is a vector of random errors.

• For each sample we create (depending only on the values of \( \varepsilon_t \)), we add a branch from the current node at stage \( t-1 \) to stage \( t \).
Random sampling

- We can generate scenarios directly from the autoregressive model:
  \[ r_t = D_0 + D_1 r_{t-1} + \cdots + D_p r_{t-p} + \varepsilon_t \]
  where \( \varepsilon_t \sim N(0, \Sigma) \) is a vector of random errors.

- For each sample we create (depending only on the values of \( \varepsilon_t \)),
  we add a branch from the current node at stage \( t-1 \) to stage \( t \).

- Many samples may be needed to approximate the distribution well.
Random sampling

- We can generate scenarios directly from the autoregressive model:
  \[ r_t = D_0 + D_1 r_{t-1} + \cdots + D_p r_{t-p} + \varepsilon_t \]
  where \( \varepsilon_t \sim N(0, \Sigma) \) is a vector of random errors.
- For each sample we create (depending only on the values of \( \varepsilon_t \)), we add a branch from the current node at stage \( t-1 \) to stage \( t \).
- Many samples may be needed to approximate the distribution well.

We want to approximate the distribution well with few samples.

**Question**

*What does it mean “approximate a distribution”?*
Random sampling

- We can generate scenarios directly from the autoregressive model:

\[ r_t = D_0 + D_1 r_{t-1} + \cdots + D_p r_{t-p} + \varepsilon_t \]

where \( \varepsilon_t \sim N(0, \Sigma) \) is a vector of random errors.

- For each sample we create (depending only on the values of \( \varepsilon_t \)), we add a branch from the current node at stage \( t - 1 \) to stage \( t \).

- Many samples may be needed to approximate the distribution well.

We want to approximate the distribution well with few samples.

---

**Question**

What does it mean “approximate a distribution”?

**Question**

What is a probability distribution uniquely defined by?
Random sampling

- We can generate scenarios directly from the autoregressive model:

\[ r_t = D_0 + D_1 r_{t-1} + \cdots + D_p r_{t-p} + \varepsilon_t \]

where \( \varepsilon_t \sim N(0, \Sigma) \) is a vector of random errors.

- For each sample we create (depending only on the values of \( \varepsilon_t \)), we add a branch from the current node at stage \( t - 1 \) to stage \( t \).

- Many samples may be needed to approximate the distribution well

We want to approximate the distribution well with few samples.

**Question**

*What does it mean “approximate a distribution”?*

**Question**

*What is a probability distribution uniquely defined by?*

**Answer:** its moments: \( E[X], E[X^2], E[X^3], \ldots \)
Fitting moments

- We now look at a sampling method that will fit all the odd moments of the distribution.
Fitting moments

- We now look at a sampling method that will fit all the odd moments of the distribution
- We then see how to fit the variance too
Fitting moments

- We now look at a sampling method that will fit all the odd moments of the distribution.
- We then see how to fit the variance too.
- With the odd moments and the variance perfectly fitted, the approximation of the distribution should be good enough for practical purposes.
Adjusted random sampling

- Assume that each node of the tree has $K = 2\ell$ branches
Adjusted random sampling

- Assume that each node of the tree has $K = 2\ell$ branches
- Instead of generating $2\ell$ samples from the autoregressive model, generate $\ell$ random samples.

This method fits all the odd moments of the distribution.
Adjusted random sampling

- Assume that each node of the tree has $K = 2\ell$ branches.
- Instead of generating $2\ell$ samples from the autoregressive model, generate $\ell$ random samples.
- Effectively this corresponds to generate $\ell$ instances of the random error vector $\varepsilon$

$\varepsilon_1, \ldots, \varepsilon_\ell$
Adjusted random sampling

- Assume that each node of the tree has $K = 2\ell$ branches.
- Instead of generating $2\ell$ samples from the autoregressive model, generate $\ell$ random samples.
- Effectively this corresponds to generate $\ell$ instances of the random error vector $\epsilon$
  \[ \epsilon_1, \ldots, \epsilon_\ell \]
- For $\ell$ branches, use the original samples.
Adjusted random sampling

• Assume that each node of the tree has $K = 2\ell$ branches
• Instead of generating $2\ell$ samples from the autoregressive model, generate $\ell$ random samples.
• Effectively this corresponds to generate $\ell$ instances of the random error vector $\varepsilon$
  \[ \varepsilon_1, \ldots, \varepsilon_\ell \]
• For $\ell$ branches, use the original samples
• For the other $\ell$, use the negative of each $\varepsilon$:
  \[ -\varepsilon_1, \ldots, -\varepsilon_\ell \]
Adjusted random sampling

- Assume that each node of the tree has $K = 2\ell$ branches.
- Instead of generating $2\ell$ samples from the autoregressive model, generate $\ell$ random samples.
- Effectively this corresponds to generate $\ell$ instances of the random error vector $\varepsilon$
  
  $$\varepsilon_1, \ldots, \varepsilon_\ell$$

- For $\ell$ branches, use the original samples.
- For the other $\ell$, use the negative of each $\varepsilon$:
  
  $$-\varepsilon_1, \ldots, -\varepsilon_\ell$$

- This method fits all the odd moments of the distribution.
Fitting the variance

• Let $\tilde{\sigma}$ the standard deviation of $\varepsilon_1, \ldots, \varepsilon_\ell$
Fitting the variance

- Let $\tilde{\sigma}$ the standard deviation of $\varepsilon_1, \ldots, \varepsilon_\ell$

- Let $\sigma$ the desired standard deviation
Fitting the variance

- Let $\tilde{\sigma}$ the standard deviation of $\varepsilon_1, \ldots, \varepsilon_\ell$

- Let $\sigma$ the desired standard deviation

- By multiplying each entry $\varepsilon_i$ by $\sigma/\tilde{\sigma}$ we obtain a set of $\varepsilon'_1, \ldots, \varepsilon'_\ell$ with the desired variance.
Tree fitting

A different approach to approximate the distribution by fitting the moments.

• Let $S_\ell$ be the value of the statistical properties (e.g., moments, variance, etc.) to be fit, for $\ell = 1, \ldots, s$.

• Let $v_k$ and $p_k$ be the vector of values on the $k$ branch from a specific node, for $k = 1, \ldots, K$.

• Let $f_\ell(v, p)$ be the mathematical expression of property $\ell$ for the discrete distribution (e.g., the mean of the vectors $v_k$).

• Assign a positive weight $w_\ell$ to each property, denoting the importance of fitting it.

• Solve the following optimization problem:

$$\min_{v, p} \sum_\ell w_\ell \left( f_\ell(v, p) - S_\ell \right)^2$$

$$\sum_k p_k = 1, p_k \geq 0$$
Tree fitting

A different approach to approximate the distribution by fitting the moments.

• Let $S_\ell$ be the value of the statistical properties (e.g., moments, variance, etc.) to be fit, for $\ell = 1, \ldots, s$
Tree fitting

A different approach to approximate the distribution by fitting the moments.

- Let $S_\ell$ be the value of the statistical properties (e.g., moments, variance, etc.) to be fit, for $\ell = 1, \ldots, s$.
- Let $v_k$ and $p_k$ be the vector of values on the $k$-branch from a specific node, for $k = 1, \ldots, K$.
Tree fitting

A different approach to approximate the distribution by fitting the moments.

• Let $S_\ell$ be the value of the statistical properties (e.g., moments, variance, etc.) to be fit, for $\ell = 1, \ldots, s$

• Let $v_k$ and $p_k$ be the vector of values on the $k$ branch from a specific node, for $k = 1, \ldots, K$

• Let $f_\ell(v, p)$ be the mathematical expression of property $\ell$ for the discrete distribution (e.g., the mean of the vectors $v_k$)
Tree fitting

A different approach to approximate the distribution by fitting the moments.

- Let $S_\ell$ be the value of the statistical properties (e.g., moments, variance, etc.) to be fit, for $\ell = 1, \ldots, s$
- Let $v_k$ and $p_k$ be the vector of values on the $k$ branch from a specific node, for $k = 1, \ldots, K$
- Let $f_\ell(v, p)$ be the mathematical expression of property $\ell$ for the discrete distribution (e.g., the mean of the vectors $v_k$)
- Assign a positive weight $w_\ell$ to each property, denoting the importance of fitting it.

\[
\min_{v, p} \sum_\ell w_\ell \left( f_\ell(v, p) - S_\ell \right)^2 \\
\sum_k p_k = 1 \quad p_k \geq 0
\]
Tree fitting

A different approach to approximate the distribution by fitting the moments.

- Let $S_\ell$ be the value of the statistical properties (e.g., moments, variance, etc.) to be fit, for $\ell = 1, \ldots, s$
- Let $v_k$ and $p_k$ be the vector of values on the $k$ branch from a specific node, for $k = 1, \ldots, K$
- Let $f_\ell(v, p)$ be the mathematical expression of property $\ell$ for the discrete distribution (e.g., the mean of the vectors $v_k$)
- Assign a positive weight $w_\ell$ to each property, denoting the importance of fitting it.
- Solve the following optimization problem:

$$\min_{v, p} \sum_{\ell} w_\ell (f_\ell(v, p) - S_\ell)^2$$

$$\sum_{k} p_k = 1$$

$$p_k \geq 0$$
Measuring risk

- Financial activities involves risk
- Until now we considered

\[ \text{variance} = \text{risk} \]

We were concerned with long term variability of an investment strategies
Measuring risk

- Financial activities involves risk
- Until now we considered variance = risk
  
  \[ \text{variance} = \text{risk} \]
  
  We were concerned with long term variability of an investment strategies
- A different idea of risk arise when consider a short(er) time horizon and rare events
Measuring risk

- Financial activities involve risk
- Until now we considered variance = risk

We were concerned with long term variability of an investment strategy.

- A different idea of risk arises when considering a short(er) time horizon and rare events.
- Financial institutions cannot be “invulnerable” to all daily horizon change in values.
Measuring risk

• Financial activities involves risk
• Until now we considered

\[
\text{variance} = \text{risk}
\]

We were concerned with long term variability of an investment strategies
• A different idea of risk arise when consider a short(er) time horizon and rare events
• Financial institutions cannot be “invulnerable” to all daily horizon change in values
• But they can measure their vulnerability in terms of losses ($) that they may incur in a single day.
• This is more a “life-or-death” kind of risk if the financial institution can not incur in a loss much greater (e.g. 3x) than one already considered very unlikely, it may not survive long
Value-at-Risk

- Let $X$ be a measure ($\$\$) of the loss of an investment in a *normal market day*, when no trading for that investment is made;

- $X$ is a random variable.

- Let $\pi$ denote the probability distribution of the daily loss $X$.

- Value-at-Risk (VaR) is a measure related to quantiles of the loss distribution $\pi$.

- It represents the predicted maximum daily loss with a specified probability level $\alpha$ (e.g., 95%):

$$\text{VaR}_\alpha(X) = \min\{\gamma : \Pr(X \geq \gamma) \leq 1 - \alpha\}$$

- Alternatively, $\text{Var}_\alpha(X)$ is the loss s.t. $\Pr(X \leq \text{Var}_\alpha(X)) = \alpha$.

- VaR summarizes the distribution of possible losses by a quantile, a single point in the domain.
Value-at-Risk

- Let $X$ be a measure ($) of the loss of an investment in a *normal market day*, when no trading for that investment is made;
- $X$ is a random variable.
Value-at-Risk

- Let $X$ be a measure ($) of the loss of an investment in a *normal market day*, when no trading for that investment is made;
- $X$ is a random variable. Let $\pi$ denote the probability distribution of the daily loss $X$

$$\pi(5M) = \Pr(X = 5M)$$
Value-at-Risk

- Let $X$ be a measure ($) of the loss of an investment in a normal market day, when no trading for that investment is made;
- $X$ is a random variable. Let $\pi$ denote the probability distribution of the daily loss $X$

$$\pi(5M) = \Pr(X = 5M)$$

- Value-at-Risk (VaR) is a measure related to quantiles of the loss distribution $\pi$. 
Value-at-Risk

- Let $X$ be a measure ($) of the loss of an investment in a normal market day, when no trading for that investment is made;
- $X$ is a random variable. Let $\pi$ denote the probability distribution of the daily loss $X$

$$\pi(5M) = \Pr(X = 5M)$$

- Value-at-Risk (VaR) is a measure related to quantiles of the loss distribution $\pi$.
- It represents the predicted maximum daily loss with a specified probability level $\alpha$ (e.g., 95%):

$$\text{VaR}_\alpha(X) = \min\{\gamma : \Pr(X \geq \gamma) \leq 1 - \alpha\}$$
Value-at-Risk

• Let $X$ be a measure ($\$\$) of the loss of an investment in a *normal market day*, when no trading for that investment is made;
• $X$ is a random variable. Let $\pi$ denote the probability distribution of the daily loss $X$

\[
\pi(5M) = \Pr(X = 5M)
\]

• Value-at-Risk (VaR) is a measure related to *quantiles of the loss distribution* $\pi$.
• It represents the predicted maximum daily loss with a specified probability level $\alpha$ (e.g., 95%):

\[
\text{VaR}_\alpha(X) = \min\{\gamma : \Pr(X \geq \gamma) \leq 1 - \alpha\}
\]

Alternatively, $\text{Var}_\alpha(X)$ is the loss s.t.

\[
\Pr(X \leq \text{Var}_\alpha(X)) = \alpha
\]

• VaR summarizes the distribution of possible losses by a quantile, a single point in the domain
When the loss distribution is continuous, \( \text{VaR}_\alpha(X) \) is simply the loss that is exceeded with probability \( \alpha \). When \( \alpha = 0.95 \), we have the 95% VaR.

Consider the example of a portfolio consisting of four investments. Suppose the portfolio has a mean return of 1% and a standard deviation of 2%. The VaR is the loss that is exceeded with probability 0.05 or 5%. The VaR can be calculated using the formula for the 95% VaR:

\[
\text{VaR}_0.95(X) = -2 \times \text{VaR} = -0.96 
\]

This means that there is a 5% chance of losing $0.96 or more.

The VaR is a useful measure of risk because it is easy to understand and it provides a single number that represents the maximum potential loss. However, it is important to note that the VaR does not take into account the potential for losses beyond the VaR.

The figure illustrates the Probability Distribution Function (PDF) of the loss distribution. The shaded area represents the 5% VaR, which is the loss that is exceeded with probability 0.05.
Interpretation

$$\text{VaR}_\alpha(X) = \min \{ \gamma : \Pr(X \geq \gamma) \leq 1 - \alpha \}$$

If my $\text{VaR}_{0.95}(X)$ is $5M$:
• is $5M$ the maximum I will lose in any day?
Interpretation

\[ \text{VaR}_\alpha(X) = \min\{\gamma : \Pr(X \geq \gamma) \leq 1 - \alpha\} \]

If my \( \text{VaR}_{0.95}(X) \) is $5M:

- is $5M the maximum I will loose in any day?
- how often do I expect to lose more than $5M?
Interpretation

$VaR_\alpha(X) = \min\{\gamma : \Pr(X \geq \gamma) \leq 1 - \alpha\}$

If my $VaR_{0.95}(X)$ is $5M$:

- is $5M$ the maximum I will loose in any day?
- how often do I expect to lose more than $5M$?
- Are there other settings in which I may lose more than $5M$?
Interpretation

\[ \text{VaR}_\alpha(X) = \min\{\gamma : \Pr(X \geq \gamma) \leq 1 - \alpha\} \]

If my \( \text{VaR}_{0.95}(X) \) is $5M:

- is $5M the maximum I will lose in any day?
- how often do I expect to lose more than $5M?
- Are there other settings in which I may lose more than $5M?
- How much shall I expect to lose if I lose more than $5M?
VaR allows to quantify how risky are normal days.

1. Be able to afford losses up to 3\times \text{VaR} (by having cash, safe investments, etc.).
2. Insure for losses over 3\times \text{VaR}. Why?

It is impossible to estimate the probability of rare events.
• VaR allows to quantify how risky are normal days
• In a normal day, it is likely that the loss won’t exceed VaR
Use / Value

- VaR allows to quantify how risky are normal days
- In a normal day, it is likely that the loss won’t exceed VaR
- On the other hand, we may lose more.
Use / Value

- VaR allows to quantify how risky are normal days
- In a normal day, it is likely that the loss won’t exceed VaR
- On the other hand, we may lose more.
- How can we use the information given by VaR to *survive*?
Use / Value

- VaR allows to quantify how risky are normal days
- In a normal day, it is likely that the loss won’t exceed VaR
- On the other hand, we may lose more.
- How can we use the information given by VaR to survive?
  1. Be able to afford losses up to $3\text{VaR}$ (by having cash, safe investments, etc.).
• VaR allows to quantify how risky are normal days
• In a normal day, it is likely that the loss won’t exceed VaR
• On the other hand, we may lose more.
• How can we use the information given by VaR to survive?

1. Be able to afford losses up to $3 \times \text{VaR}$ (by having cash, safe investments, etc.).

2. Insure for losses over $3 \times \text{VaR}$. Why?
• VaR allows to quantify how risky are normal days
• In a normal day, it is likely that the loss won’t exceed VaR
• On the other hand, we may lose more.
• How can we use the information given by VaR to *survive*?
  1. Be able to afford losses up to $3\text{VaR}$ (by having cash, safe investments, etc.).
  2. Insure for losses over $3\text{VaR}$. Why?
    - It is impossible to estimate the probability of rare events.
Lack of subadditivity

- VaR is somewhat intuitive, but it lacks a desirable property for risk measures: subadditivity
- Risk measures should respect the idea that “diversification reduces risk”
Lack of subadditivity

- VaR is somewhat intuitive, but it lacks a desirable property for risk measures: *subadditivity*
- Risk measures should respect the idea that “diversification reduces risk”
- I.e., the total risk of two different investment portfolios should not exceed the sum of the individual risks
Lack of subadditivity

• VaR is somewhat intuitive, but it lacks a desirable property for risk measures: *subadditivity*
• Risk measures should respect the idea that “diversification reduces risk”
• I.e., the total risk of two different investment portfolios should not exceed the sum of the individual risks
• I.e., for a risk measure $f$, we should have

$$f(x_1 + x_2) \leq f(x_1) + f(x_2), \forall x_1, x_2$$
Example

- Two *independent* investment strategies, $A$ and $B$
Example

- Two *independent* investment strategies, $A$ and $B$
- Each of $A$ and $B$ returns:
  - gain of $1$ (loss of $-1$) with probability $0.96$
  - loss of $2$ with probability $0.04$
- $\text{VaR}_{0.95}(A) = \text{VaR}_{0.95}(B) =$
Example

- Two *independent* investment strategies, $A$ and $B$
- Each of $A$ and $B$ returns:
  - gain of $1$ (loss of $-1$) with probability 0.96
  - loss of $2$ with probability 0.04
- $\text{VaR}_{0.95}(A) = \text{VaR}_{0.95}(B) = \min\{\gamma : \Pr(A \geq \gamma) \leq 0.05\} =$
Example

- Two **independent** investment strategies, \( A \) and \( B \)
- Each of \( A \) and \( B \) returns:
  - gain of $1 (loss of $−1) with probability 0.96
  - loss of $2 with probability 0.04
- \( \text{VaR}_{0.95}(A) = \text{VaR}_{0.95}(B) = \min\{\gamma : \Pr(A \geq \gamma) \leq 0.05\} = -1 \)
Example

- Two independent investment strategies, $A$ and $B$
- Each of $A$ and $B$ returns:
  - gain of $1$ (loss of $-1$) with probability $0.96$
  - loss of $2$ with probability $0.04$
- $\text{VaR}_{0.95}(A) = \text{VaR}_{0.95}(B) = \min\{\gamma : \Pr(A \geq \gamma) \leq 0.05\} = -1$
- Consider now the investment strategy $= A + B$
- $C$ returns:
Example

- Two *independent* investment strategies, $A$ and $B$
- Each of $A$ and $B$ returns:
  - gain of $1$ (loss of $-1$) with probability $0.96$
  - loss of $2$ with probability $0.04$
- $\text{VaR}_{0.95}(A) = \text{VaR}_{0.95}(B) = \min\{\gamma : \Pr(A \geq \gamma) \leq 0.05\} = -1$
- Consider now the investment strategy $= A + B$
- $C$ returns:
  - loss of $-2$ with prob. $0.96 \times 0.96 = 0.9216$
Example

- Two independent investment strategies, $A$ and $B$
- Each of $A$ and $B$ returns:
  - gain of $1$ (loss of $-1$) with probability $0.96$
  - loss of $2$ with probability $0.04$
- $\text{VaR}_{0.95}(A) = \text{VaR}_{0.95}(B) = \min\{\gamma : \Pr(A \geq \gamma) \leq 0.05\} = -1$
- Consider now the investment strategy $= A + B$
- $C$ returns:
  - loss of $-2$ with prob. $0.96 \times 0.96 = 0.9216$
  - loss of $1$ with prob. $2 \times 0.96 \times 0.0 = 0.0768$
Example

- Two *independent* investment strategies, $A$ and $B$
- Each of $A$ and $B$ returns:
  - gain of $1$ (loss of $-1$) with probability $0.96$
  - loss of $2$ with probability $0.04$
- $\text{VaR}_{0.95}(A) = \text{VaR}_{0.95}(B) = \min\{\gamma : \Pr(A \geq \gamma) \leq 0.05\} = -1$
- Consider now the investment strategy $C = A + B$
- $C$ returns:
  - loss of $-2$ with prob. $0.96 \times 0.96 = 0.9216$
  - loss of $1$ with prob. $2 \times 0.96 \times 0.0 = 0.0768$
  - loss of $4$ with prob. $0.04 \times 0.04 = 0.0016$
Example

• Two independent investment strategies, $A$ and $B$

• Each of $A$ and $B$ returns:
  • gain of $1$ (loss of $-1$) with probability 0.96
  • loss of $2$ with probability 0.04

• $VaR_{0.95}(A) = VaR_{0.95}(B) = \min\{\gamma : \Pr(A \geq \gamma) \leq 0.05\} = -1$

• Consider now the investment strategy $= A + B$

• $C$ returns:
  • loss of $-2$ with prob. $0.96 \times 0.96 = 0.9216$
  • loss of $1$ with prob. $2 \times 0.96 \times 0.0 = 0.0768$
  • loss of $4$ with prob. $0.04 \times 0.04 = 0.0016$

• Then $VaR_{0.95}(A + B) =$
Example

- Two independent investment strategies, $A$ and $B$
- Each of $A$ and $B$ returns:
  - gain of $1$ (loss of $-1$) with probability $0.96$
  - loss of $2$ with probability $0.04$
- $\text{VaR}_{0.95}(A) = \text{VaR}_{0.95}(B) = \min\{\gamma : \Pr(A \geq \gamma) \leq 0.05\} = -1$
- Consider now the investment strategy $= A + B$
- $C$ returns:
  - loss of $-2$ with prob. $0.96 \times 0.96 = 0.9216$
  - loss of $1$ with prob. $2 \times 0.96 \times 0.0 = 0.0768$
  - loss of $4$ with prob. $0.04 \times 0.04 = 0.0016$
- Then $\text{VaR}_{0.95}(A + B) = 1$
Example

- Two independent investment strategies, $A$ and $B$
- Each of $A$ and $B$ returns:
  - gain of $1$ (loss of $-1$) with probability 0.96
  - loss of $2$ with probability 0.04
- $\text{VaR}_{0.95}(A) = \text{VaR}_{0.95}(B) = \min\{\gamma : \Pr(A \geq \gamma) \leq 0.05\} = -1$
- Consider now the investment strategy $= A + B$
- $C$ returns:
  - loss of $-2$ with prob. $0.96 \times 0.96 = 0.9216$
  - loss of $1$ with prob. $2 \times 0.96 \times 0.0 = 0.0768$
  - loss of $4$ with prob. $0.04 \times 0.04 = 0.0016$
- Then $\text{VaR}_{0.95}(A + B) = 1 > \text{VaR}_{0.95}(A) + \text{VaR}_{0.95}(B) = -2$
• Computing VaR by generating scenarios is difficult. It is a non-smooth, non-convex function of the positions in the investment portfolio. This results in multiple local optimals that hinder the process.
Other criticisms of VaR

• Computing VaR by generating scenarios is difficult. It is a non-smooth, non-convex function of the positions in the investment portfolio. This results in multiple local optimals that hinder the process.

• It pays no attention to the magnitude of the losses beyond the VaR value.
Other criticisms of VaR

- Computing VaR by generating scenarios is difficult. It is a non-smooth, non-convex function of the positions in the investment portfolio. This results in multiple local optimals that hinder the process.
- It pays no attention to the magnitude of the losses beyond the VaR value
- There are other issues with VaR, both mathematical/statistical and “philosophical”. 

Other criticisms of VaR

- Computing VaR by generating scenarios is difficult. It is a non-smooth, non-convex function of the positions in the investment portfolio. This results in multiple local optimals that hinder the process.
- It pays no attention to the magnitude of the losses beyond the VaR value.
- There are other issues with VaR, both mathematical/statistical and “philosophical”. They arise mostly because people do not understand probabilities and misinterpret VaR (more on this later).
Conditional Value at Risk (CVaR) addresses one of the criticisms.
Conditional Value-at-Risk

- Conditional Value at Risk (CVaR) addresses one of the criticisms
- It answers the question:

<table>
<thead>
<tr>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>What is the expected loss, given that the loss exceeds VaR?</td>
</tr>
</tbody>
</table>
Conditional Value-at-Risk

- Conditional Value at Risk (CVaR) addresses one of the criticisms.
- It answers the question:

  **Question**

  *What is the expected loss, given that the loss exceeds VaR?*

- It can be computed using LP when the loss function is linear in the portfolio positions.
Conditional Value-at-Risk

• $x$: the vector of portfolio positions.
  $x \in X$ ($X$: set of feasible portfolios)
Conditional Value-at-Risk

- $x$: the vector of portfolio positions.
  \[ x \in X \quad (X: \text{set of feasible portfolios}) \]
- $y$: the vector of random events.
  \[ p(y): \text{probability density function of } y \]
Conditional Value-at-Risk

- \( x \): the vector of portfolio positions.
  \( x \in X \) (\( X \): set of feasible portfolios)
- \( y \): the vector of random events.
  \( p(y) \): probability density function of \( y \)
- \( f(x, y) \): loss function for portfolio \( x \) and realization \( y \)
Conditional Value-at-Risk

- $x$: the vector of portfolio positions.
  \[ x \in X \quad (X: \text{set of feasible portfolios}) \]
- $y$: the vector of random events.
  \[ p(y): \text{probability density function of } y \]
- $f(x, y)$: loss function for portfolio $x$ and realization $y$
- For a fixed $x$, the cumulative distribution function of the loss associated to $x$ is:
  \[
  \Psi(x, \gamma) = \int_{f(x,y)<\gamma} p(y)dy
  \]
Conditional Value-at-Risk

- \( x \): the vector of portfolio positions.
  \[ x \in X \quad (X: \text{set of feasible portfolios}) \]
- \( y \): the vector of random events.
  \[ p(y) \text{: probability density function of } y \]
- \( f(x, y) \): loss function for portfolio \( x \) and realization \( y \)
- For a fixed \( x \), the cumulative distribution function of the loss associated to \( x \) is:
  \[ \Psi(x, \gamma) = \int_{f(x,y)<\gamma} p(y) dy \]
  \[ (\text{VaR}_\alpha(x) = \min\{\gamma \in \mathbb{R} : \Psi(x, \gamma) \geq \alpha\}) \]
Conditional Value-at-Risk

- $x$: the vector of portfolio positions.
  \[ x \in X \] ($X$: set of feasible portfolios)
- $y$: the vector of random events.
  \[ p(y): \text{probability density function of } y \]
- $f(x, y)$: loss function for portfolio $x$ and realization $y$
- For a fixed $x$, the cumulative distribution function of the loss associated to $x$ is:
  \[ \Psi(x, \gamma) = \int_{f(x, y) < \gamma} p(y) dy \]
  \[ (\text{VaR}_\alpha(x) = \min\{\gamma \in \mathbb{R} : \Psi(x, \gamma) \geq \alpha\}) \]
Conditional Value-at-Risk

• the $\alpha - CVaR$ of $x$ is

$$CVaR_\alpha(x) = \frac{1}{1 - \alpha} \int_{f(x,y) \geq VaR_\alpha(x)} f(x, y)p(y)dy$$
Conditional Value-at-Risk

- the $\alpha - CVaR$ of $x$ is

$$CVaR_\alpha(x) = \frac{1}{1 - \alpha} \int_{f(x,y) \geq VaR_\alpha(x)} f(x,y)p(y)dy$$

For a discrete probability distribution:

$$CVaR_\alpha(x) = \frac{1}{1 - \alpha} \sum_{j: f(x,y_j) \geq VaR_\alpha(x)} p_j f(x,y_j)$$
Conditional Value-at-Risk

- the $\alpha - CVaR$ of $x$ is

$$CVaR_\alpha(x) = \frac{1}{1 - \alpha} \int_{f(x,y) \geq VaR_\alpha(x)} f(x,y)p(y)dy$$

For a discrete probability distribution:

$$CVaR_\alpha(x) = \frac{1}{1 - \alpha} \sum_{j : f(x,y_j) \geq VaR_\alpha(x)} p_j f(x,y_j)$$

- CVaR is the conditional expectation of the loss given that the loss is greater than VaR.
CVaR versus VaR

Fact

\[ CVaR_\alpha(x) \geq VaR_\alpha(x) \]
CVaR versus VaR

Fact

\[ CVaR_\alpha(x) \geq VaR_\alpha(x) \]

Proof.

\[
CVaR_\alpha(x) = \frac{1}{\alpha} \int_{f(x,y) \geq VaR_\alpha(x)} f(x,y)p(y)dy \\
\geq \frac{1}{1 - \alpha} \int_{f(x,y) \geq VaR_\alpha(x)} VaR_\alpha(x)p(y)dy \\
= \frac{VaR_\alpha(x)}{1 - \alpha} \int_{f(x,y) \geq VaR_\alpha(x)} p(y)dy \\
\geq VaR_\alpha(x)
\]
Example

- Let $j$ be a uniform random variable in $[-75, 24]$. 
  \[ \Pr(j = -75) = \Pr(j = -74) = \cdots = \Pr(j = 24) = 0.01 \]
Example

- Let \( j \) be a uniform random variable in \([-75, 24]\).
  \[
  \Pr(j = -75) = \Pr(j = -74) = \cdots = \Pr(j = 24) = 0.01
  \]

- Fix \( x \) and let \( f(x, y) = y \)
Example

• Let $j$ be a uniform random variable in $[-75, 24]$.
  \[ \Pr(j = -75) = \Pr(j = -74) = \cdots = \Pr(j = 24) = 0.01 \]
• Fix $x$ and let $f(x, y) = y$
• $\text{VaR}_{0.95}(x) =$
Example

• Let $j$ be a uniform random variable in $[-75, 24]$.
  \[
  \Pr(j = -75) = \Pr(j = -74) = \cdots = \Pr(j = 24) = 0.01
  \]
• Fix $x$ and let $f(x, y) = y$
• $\text{VaR}_{0.95}(x) = 20$ as the loss is 20 or more with probability 5%
Example

• Let \( j \) be a uniform random variable in \([-75, 24]\).
  \[
  \Pr(j = -75) = \Pr(j = -74) = \cdots = \Pr(j = 24) = 0.01
  \]

• Fix \( x \) and let \( f(x, y) = y \)

• \( \text{VaR}_{0.95}(x) = 20 \) as the loss is 20 or more with probability 5%

• Since

\[
\text{CVaR}_\alpha(x) = \frac{1}{1 - \alpha} \sum_{j: f(x, y_j) \geq \text{VaR}_\alpha(x)} p_j f(x, y_j)
\]

We have

\[
\text{CVaR}_{0.95}(x) = \frac{1}{1 - 0.95} \sum_{j: f(x, y_j) \geq \text{VaR}_{0.95}(x)} p_j f(x, y_j)
\]

\[
= \frac{1}{0.05} (20 + 21 + 22 + 23 + 24) \times 0.01 = 22 (> 20)
\]
Minimizing CVaR

- The definition of CVaR involves the VaR function explicitly

**Question**

*We want to find the portfolio allocation $x \in X$ with minimum CVaR.*
Minimizing CVaR

- The definition of CVaR involves the VaR function explicitly.

**Question**

We want to find the portfolio allocation \( x \in X \) with minimum CVaR.

*It is difficult to optimize VaR, is it difficult to optimize CVaR?*
Minimizing CVaR

- The definition of CVaR involves the VaR function explicitly

**Question**

We want to find the portfolio allocation $x \in X$ with minimum CVaR.

*It is difficult to optimize VaR, is it difficult to optimize CVaR?*

- Auxiliary function:

\[
F_\alpha(x, \gamma) = \gamma + \frac{1}{1 - \alpha} \int_{f(x,y) \geq \gamma} (f(x,y) - \gamma)p(y)dy
\]
Minimizing CVaR

• The definition of CVaR involves the VaR function explicitly

**Question**

*We want to find the portfolio allocation* \( x \in X \) *with minimum CVaR.*

*It is difficult to optimize VaR, is it difficult to optimize CVaR?*

• Auxiliary function:

\[
F_\alpha(x, \gamma) = \gamma + \frac{1}{1 - \alpha} \int_{f(x, y) \geq \gamma} (f(x, y) - \gamma)p(y)dy
\]

• \( F_\alpha(x, \gamma) \) has the following good properties:
  • \( F_\alpha(x, \gamma) \) is a convex function of \( \gamma \)
Minimizing CVaR

• The definition of CVaR involves the VaR function explicitly

**Question**

We want to find the portfolio allocation \( x \in X \) with minimum CVaR.

*It is difficult to optimize VaR, is it difficult to optimize CVaR?*

• Auxiliary function:

\[
F_\alpha(x, \gamma) = \gamma + \frac{1}{1 - \alpha} \int_{f(x,y) \geq \gamma} (f(x,y) - \gamma)p(y)dy
\]

• \( F_\alpha(x, \gamma) \) has the following good properties:
  • \( F_\alpha(x, \gamma) \) is a convex function of \( \gamma \)
  • \( \text{VaR}_\alpha(x) \) is a minimizer over \( \gamma \) of \( F_\alpha(x, \gamma) \)
Minimizing CVaR

- The definition of CVaR involves the VaR function explicitly.

**Question**

We want to find the portfolio allocation \( x \in X \) with minimum CVaR.

*It is difficult to optimize VaR, is it difficult to optimize CVaR?*

- Auxiliary function:

\[
F_\alpha(x, \gamma) = \gamma + \frac{1}{1 - \alpha} \int_{f(x,y) \geq \gamma} (f(x,y) - \gamma)p(y)dy
\]

- \( F_\alpha(x, \gamma) \) has the following good properties:
  - \( F_\alpha(x, \gamma) \) is a convex function of \( \gamma \)
  - \( \text{VaR}_\alpha(x) \) is a minimizer over \( \gamma \) of \( F_\alpha(x, \gamma) \)
  - The minimum value over \( \gamma \) of \( F_\alpha(x, \gamma) \) is \( \text{CVaR}_\alpha(x) \)
Minimizing CVaR

• The definition of CVaR involves the VaR function explicitly

Question

We want to find the portfolio allocation $x \in X$ with minimum CVaR.

It is difficult to optimize VaR, is it difficult to optimize CVaR?

• Auxiliary function:

$$F_\alpha(x, \gamma) = \gamma + \frac{1}{1 - \alpha} \int_{f(x,y) \geq \gamma} (f(x,y) - \gamma)p(y)dy$$

• $F_\alpha(x, \gamma)$ has the following good properties:
  • $F_\alpha(x, \gamma)$ is a convex function of $\gamma$
  • $\text{VaR}_\alpha(x)$ is a minimizer over $\gamma$ of $F_\alpha(x, \gamma)$
  • The minimum value over $\gamma$ of $F_\alpha(x, \gamma)$ is $\text{CVaR}_\alpha(x)$
Minimizing CVaR

• The definition of CVaR involves the VaR function explicitly

Question

We want to find the portfolio allocation $x \in X$ with minimum CVaR.

It is difficult to optimize VaR, is it difficult to optimize CVaR?

• Auxiliary function:

$$F_\alpha(x, \gamma) = \gamma + \frac{1}{1 - \alpha} \int_{f(x, y) \geq \gamma} (f(x, y) - \gamma)p(y)dy$$

• $F_\alpha(x, \gamma)$ has the following good properties:
  • $F_\alpha(x, \gamma)$ is a convex function of $\gamma$
  • $VaR_\alpha(x)$ is a minimizer over $\gamma$ of $F_\alpha(x, \gamma)$
  • The minimum value over $\gamma$ of $F_\alpha(x, \gamma)$ is $CVaR_\alpha(x)$

(Proof: homework)
Minimizing CVaR

- The definition of CVaR involves the VaR function explicitly

**Question**

*We want to find the portfolio allocation \( x \in X \) with minimum CVaR.*

*It is difficult to optimize VaR, is it difficult to optimize CVaR?*

- Auxiliary function:

\[
F_\alpha(x, \gamma) = \gamma + \frac{1}{1 - \alpha} \int_{f(x,y) \geq \gamma} (f(x,y) - \gamma)p(y)dy
\]

- \( F_\alpha(x, \gamma) \) has the following good properties:
  - \( F_\alpha(x, \gamma) \) is a convex function of \( \gamma \)
  - \( \text{VaR}_\alpha(x) \) is a minimizer over \( \gamma \) of \( F_\alpha(x, \gamma) \)
  - The minimum value over \( \gamma \) of \( F_\alpha(x, \gamma) \) is \( \text{CVaR}_\alpha(x) \)

(Proof: homework)
\[ F_\alpha(x, \gamma) = \gamma + \frac{1}{1 - \alpha} \int_{f(x,y) \geq \gamma} (f(x,y) - \gamma)p(y)dy \]

- To minimize \( CVaR_\alpha(x) \) over \( x \), we need to minimize \( F_\alpha(x, \gamma) \) w.r.t. \( x \) and \( \gamma \):

\[
\min_{x \in X} CVaR_\alpha(x) = \min_{x \in X, \gamma} F_\alpha(x, \gamma)
\]
\[ F_\alpha(x, \gamma) = \gamma + \frac{1}{1 - \alpha} \int_{f(x,y) \geq \gamma} (f(x,y) - \gamma)p(y)dy \]

- To minimize \( CVaR_\alpha(x) \) over \( x \), we need to minimize \( F_\alpha(x, \gamma) \) w.r.t. \( x \) and \( \gamma \):

\[
\min_{x \in X} CVaR_\alpha(x) = \min_{x \in X, \gamma} F_\alpha(x, \gamma)
\]

No need to compute VaR first.
\[ F_\alpha(x, \gamma) = \gamma + \frac{1}{1 - \alpha} \int_{f(x,y) \geq \gamma} (f(x,y) - \gamma)p(y)dy \]

- To minimize \( CVaR_\alpha(x) \) over \( x \), we need to minimize \( F_\alpha(x, \gamma) \) w.r.t. \( x \) and \( \gamma \):

\[
\min_{x \in X} CVaR_\alpha(x) = \min_{x \in X, \gamma} F_\alpha(x, \gamma)
\]

No need to compute VaR first.

- If \( f(x, y) \) is a convex (linear) function of \( x \), then \( F_\alpha(x, \gamma) \) is also a convex (linear) function of \( x \).
\[ F_{\alpha}(x, \gamma) = \gamma + \frac{1}{1 - \alpha} \int_{f(x, y) \geq \gamma} (f(x, y) - \gamma) p(y) dy \]

- To minimize \( CVaR_{\alpha}(x) \) over \( x \), we need to minimize \( F_{\alpha}(x, \gamma) \) w.r.t. \( x \) and \( \gamma \):

\[
\min_{x \in X} CVaR_{\alpha}(x) = \min_{x \in X, \gamma} F_{\alpha}(x, \gamma)
\]

No need to compute VaR first.

- If \( f(x, y) \) is a convex (linear) function of \( x \), then \( F_{\alpha}(x, \gamma) \) is also a convex (linear) function of \( x \)

- If the set \( X \) of feasible portfolios is convex, then we have a smooth convex optimization problem, which we know how to solve!
In practice. . .

• In practice, we may not have the joint density function \( p(y) \) of the random events
In practice...

• In practice, we may not have the joint density function \( p(y) \) of the random events

• We have a collection of scenarios \( y_s, s = 1, \ldots, S \), obtained from historical data or by simulations.
In practice... 

• In practice, we may not have the joint density function $p(y)$ of the random events.

• We have a collection of scenarios $y_s$, $s = 1, \ldots, S$, obtained from historical data or by simulations.

• We assume all the scenarios have the same probability (empirical distribution).
In practice...

- In practice, we may not have the joint density function $p(y)$ of the random events.
- We have a collection of scenarios $y_s, s = 1, \ldots, S$, obtained from historical data or by simulations.
- We assume all the scenarios have the same probability (empirical distribution).
- We can approximate $F_\alpha(x, \gamma)$ as:

$$\tilde{F}_\alpha(x, \gamma) = \gamma + \frac{1}{(1 - \alpha)S} \sum_{s=1}^{S} \max\{f(x, y) - \gamma, 0\}$$
In practice...

- In practice, we may not have the joint density function $p(y)$ of the random events.
- We have a collection of scenarios $y_s$, $s = 1, \ldots, S$, obtained from historical data or by simulations.
- We assume all the scenarios have the same probability (empirical distribution).
- We can approximate $F_\alpha(x, \gamma)$ as:
  $$\tilde{F}_\alpha(x, \gamma) = \gamma + \frac{1}{(1 - \alpha)S} \sum_{s=1}^{S} \max\{f(x, y) - \gamma, 0\}$$

- We want to find
  $$\min_{x \in X, \gamma} \gamma + \frac{1}{(1 - \alpha)S} \sum_{s=1}^{S} \max\{f(x, y) - \gamma, 0\}$$
Solving the optimization problem

\[
\min_{x \in X, \gamma} \gamma + \frac{1}{(1 - \alpha)S} \sum_{s=1}^{S} \max\{f(x, y) - \gamma, 0\}
\]
Solving the optimization problem

\[
\min_{x \in X, \gamma} \gamma + \frac{1}{(1 - \alpha)S} \sum_{s=1}^{S} \max\{f(x, y) - \gamma, 0\}
\]

We introduce artificial variables \(z_s\) and constraints on them to remove the \(\max\) from the objective function:
Solving the optimization problem

\[
\min_{x \in X, \gamma} \gamma + \frac{1}{(1 - \alpha)S} \sum_{s=1}^{S} \max\{f(x, y) - \gamma), 0\}
\]

We introduce artificial variables \(z_s\) and constraints on them to remove the \(\max\) from the objective function:

\[
\min_{x, z, \gamma} \gamma \frac{1}{(1 - \alpha)S} \sum_{s=1}^{S} z_s
\]

s.t. \(z_s \geq 0, s = 1, \ldots, S\)

\(z_s \geq f(x, y_s) - \gamma, s = 1, \ldots, S\)

\(x \in X\)
Solving the optimization problem

\[
\min_{x \in X, \gamma} \gamma + \frac{1}{1 - \alpha} \sum_{s=1}^{S} \max\{f(x, y) - \gamma, 0\}
\]

We introduce artificial variables \(z_s\) and constraints on them to remove the \(\max\) from the objective function:

\[
\min_{x, z, \gamma} \gamma \frac{1}{1 - \alpha} \sum_{s=1}^{S} z_s
\]

s.t. \(z_s \geq 0, s = 1, \ldots, S\)

\(z_s \geq f(x, y_s) - \gamma, s = 1, \ldots, S\)

\(x \in X\)

If \(f(x, y)\) is linear in \(x\), we have all linear constraints and a linear objective, hence we have an LP that we can solve efficiently.
Other problems in risk management

- Often, investments managers want to optimize a performance measure (e.g., expected returns) under risk constraints.
- CVaR can be used in this setting:

\[
\begin{align*}
\max_x & \quad \mu^T x \\
\text{s.t.} & \quad \text{CVaR}_{\alpha_j}(x) \leq U_{\alpha_j}, j = 1, \ldots, J \\
& \quad x \in X
\end{align*}
\]
Other problems in risk management

- Often, investments managers want to optimize a performance measure (e.g., expected returns) under risk constraints. CVaR can be used in this setting:

\[
\max_{x} \mu^T x \\
\text{s.t. } CVaR_{\alpha_j}(x) \leq U_{\alpha_j}, j = 1, \ldots, J \\
x \in X
\]

- We can use historical/simulated scenarios, then introduce auxiliary variables as before and obtain a LP formulation.
Criticism to CVaR

- Is CVaR even a useful measure?
Criticism to CVaR

- Is CVaR even a useful measure?
- The losses above VaR are, by definition, rare events
Criticism to CVaR

- Is CVaR even a useful measure?
- The losses above VaR are, by definition, rare events
- We are in the tail of the distribution, which is often fat, but we can’t estimate the probabilities in the tail.
Criticism to CVaR

- Is CVaR even a useful measure?
- The losses above VaR are, by definition, *rare events*
- We are in the tail of the distribution, which is often *fat*, but we can’t estimate the probabilities in the tail.
- How can we compute an expectation, even approximate, if we do not have the probabilities?
Black Swans

Figure: Nassim Nicholas Taleb – from Wikipedia

- Books: Fooled by Randomness, The Black Swan, Antifragile,
Black Swans

Figure: Nassim Nicholas Taleb – from Wikipedia

- Books: Fooled by Randomness, The Black Swan, Antifragile,
- Black Swans: extremely rare events with extremely large impact
Black Swans

Figure: Nassim Nicholas Taleb – from Wikipedia

- Books: Fooled by Randomness, The Black Swan, Antifragile,
- Black Swans: extremely rare events with extremely large impact
- They are actually the events that drive history
Black Swans

Books: Fooled by Randomness, The Black Swan, Antifragile,
Black Swans: extremely rare events with extremely large impact
They are actually the events that drive history
As Black Swans are rare events, they are impossible to predict;

Figure: Nassim Nicholas Taleb – from Wikipedia
Black Swans

- Books: Fooled by Randomness, The Black Swan, Antifragile,
- Black Swans: extremely rare events with extremely large impact
- They are actually the events that drive history
- As Black Swans are rare events, they are impossible to predict;
- In hindsight, people wrongly claim that they could have been predicted;

Figure: Nassim Nicholas Taleb – from Wikipedia
**Black Swans**

- Books: Fooled by Randomness, The Black Swan, Antifragile,
- Black Swans: extremely rare events with extremely large impact
- They are actually the events that drive history
- As Black Swans are rare events, they are impossible to predict;
- In hindsight, people wrongly claim that they could have been predicted;
- Rather than try to predict them, build systems that are
  - robust to negative ones, and
  - able to exploit positive ones