Model Uncertainty in Energy Optimization

Georg Ch. Pflug/A. Pichler/D. Wozabal/B. Analui

June 1, 2017

Optimal decision making under uncertainty

We consider two types of instances in optimal decision making in energy

- Pricing of contracts
- Optimal management of resources

In both cases a stochastic model for the uncertain parameters is needed.

Pricing principles

- *Superreplication pricing:* The minimal price is found, which allows the seller to hedge all risks away
- *Acceptability pricing:* The minimal price is found, which allows the seller to hedge the contract is such a way, that the risks are acceptable.
- *Indifference pricing:* The risk limit for the accaptability price is found by considering the risk exposure of the seller before he/she concludes the contract.
- *Ambiguity pricing:* The model risk is included in the pricing algorithm.

Acceptability pricing for a fixed energy delivery contract

C_t	payments (cash-inflow) to the contract seller
$\mathcal{J} = \{0, \ldots, J\}$	energy forms (electricity: $j = 0$, gas, oil, water)
$S^e_{t,j}$	spot prices
$0 \leq x^e_{t,j} \leq \bar{x}^e_j$	storage and constraints (for electricity $\bar{x}^e_j = 0$)
$x^f_{t,0}$	cash with an interest yield of $r_f > 0$
$y^e_t = (y^e_{t,0}, \ldots, y_{t,J})$	amount of energy bought or sold
$d_{t,j} \geq 0$	random inflows (solar, wind, water)
$z^e_{t,ij}$	production of energy i out of energy j
$\underline{z}_{t,ij} \leq z^e_{t,ij} \leq \bar{z}_{t,ij}$	production limit
η_{ij}	efficiencies of conversion
$\gamma_{t,ij}$	cost factors
$S^f_{t,i}$	financial assets paying cash flows $C^f_{t,i}$
$D_t(t,j)$	delivery of energy j in period $[t, t+1)$ in MWh to the

Equations and constraints

Initialization of energy storages (except electricity)

$$x^e_{0,j} \leq x^*_j + y^e_{0,j} + d_{0,j} \tag{1}$$

and

$$x^e_{t,j} \leq x^e_{t-1,j} + y^e_{t,j} + \sum_{i=0}^{J} \eta_{ij} z^e_{t-1,ij} - \sum_{i=1}^{J} z^e_{t-1,ji} + d_{t,j} - D_{t,j}. \tag{2}$$

For electricity

$$0 = y^e_{0,0} + d_{0,0} \tag{3}$$

$$0 = y^e_{t,0} + \sum_{i=1}^{J} \eta_{i0} z^e_{t-1,i0} + d_{t,0} - D_{t,0}. \tag{4}$$

Only energy stored at the beginning of a period can be used for conversion during the period:

$$\sum_{j=1}^{J} z^e_{t,ij} \leq x^e_{t,i}. \tag{5}$$

Initial cash-account

$$x^f_{0,0} \leq w - \sum_{j=0}^{J} S^e_{0,j} y^e_{0,j} - \sum_{i=0}^{I} S^f_{0,i} x^f_{0,i}. \tag{6}$$

and later

$$\begin{aligned} x^f_{t,0} &\leq (1+r_f)x^f_{t-1,0} \\ &- \sum_{j=0}^{J} S^e_{t,j} y^e_{t,j} - \sum_{i=1}^{I} S^f_{t,i}(x^f_{t,i} - x^f_{t-1,i}) + \sum_{i=1}^{I} C^f_{t,i} + C_t \\ &- \sum_{i=0}^{J}\sum_{j=0}^{J} \gamma_{t,ij} z_{t,ij} - \sum_{j=1}^{J} \zeta_j \frac{(x^e_{t,j} + x^e_{t-1,j})}{2} \end{aligned} \tag{7}$$

The terminal inequality ensures that the final asset value is nonnegative

$$x^f_{T,0} + \sum_{j=1}^{J} S^e_{T,j} x^e_{t,j} + \sum_{i=1}^{I} S^f_{T,i} x^f_{T,i} \geq 0. \tag{8}$$

Superreplication pricing as an optimization problem

The (superreplication) price is the minimal value of the following optimization problem:

$$\left\| \begin{array}{l} \text{Minimize (in } x^e, x^f, y, z \text{ and } w) : \; w \\ \text{subject to all constraints} \\ x_t^e, x_t^f, y_t, z_t \text{ are non-anticipative.} \end{array} \right. \tag{9}$$

Acceptability pricing as a stochastic optimization problem

The optimization problem for acceptability pricing is a modification and can be written as

$$
\left\| \begin{array}{l} \text{Minimize (in } x^e, x^f, y, z \text{ and } w) : \ w \\ \text{subject to the given constraints} \\ \mathcal{A}(x^f_{T,0} + \sum_{j=1}^{J} S^e_{T,j} x^e_{t,j} + \sum_{i=1}^{I} S^f_{T,i} x^f_{T,i}) \geq 0 \\ x^e_t, x^f_t, y_t, z_t \text{ are non-anticipative.} \end{array} \right. \qquad (10)
$$

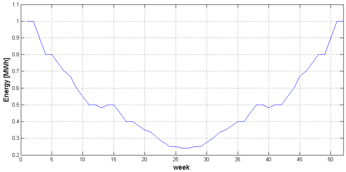
If $\mathcal{A} = \mathit{essinf}$ we get superreplication price, otherwise we get an acceptability price.

An example

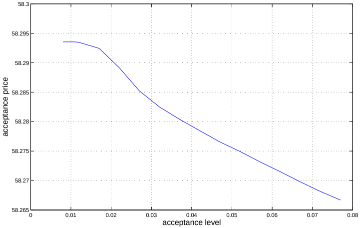
We consider a planning horizon of one year (52 weeks). Electricity spot prices are modeled by geometric Brownian motion with jumps (GBMJ), estimated from EEX Phelix hourly electricity prices (hourly, 09/2008-12/2011, Bloomberg). The pricing model was discretized in time and space by generating a tree process, generated from the GBMJ model.

The hedging opportunities are represented by four futures contracts, related to the quarters of the year, i.e. each of the futures delivers a constant amount of electric energy during one of the quarters.

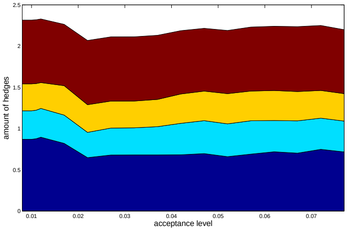
By solving the stochastic optimization problem with the average value-at-risk ($\mathbb{A}$V@R$_\alpha$) as acceptability functional, the acceptability price is calculated for a pure trader meaning that only wholesale base quarter future contracts can be used for hedging for different values of the $\mathbb{A}$V@R-parameter α.



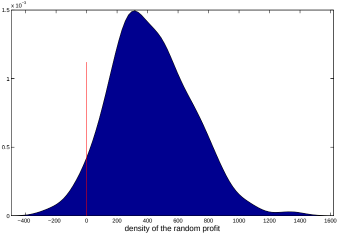
Acceptability pricing: delivery pattern D_t over 52 weeks.



Acceptability pricing: The price of 1 MWh as a function of the acceptance level α.



Acceptability pricing: optimal hedges as a function of the acceptance level α.



Acceptability pricing: density of the profit variable

Ambiguity

Traditionally, optimal decision making under uncertainty is done two steps:

- Step 1: Estimation of a probability model for the random scenarios
- Step 2: Finding the best decision given the estimated model

According to Ellsberg (1961) we face here two types of non-determinism:
Uncertainty: the probabilistic model is known, but the realizations of the random variables are unknown ("aleatoric uncertainty")
Ambiguity: the probability model itself is not fully known ("epistemic uncertainty").
Ambiguity sets $\mathcal{P}$: A family of probability models $\mathcal{P}$ which are all plausible models for the reality and we are uncertain about which concrete $P \in \mathcal{P}$ is the true one.

Problem formulation: Ambiguity

Let the basic problem be

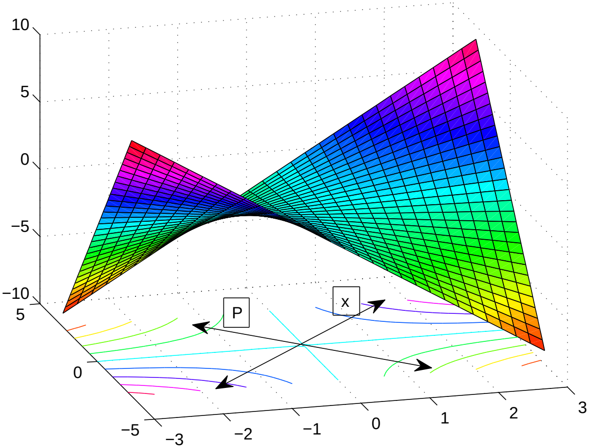
$$\min \left\{ \mathbb{E}_{\hat{P}}[Q(x,\xi)] \,:\, x \in \mathbb{X} \right\}$$

and let $\mathcal{P}$ be the ambiguity set. Then the ambiguity problem is

$$\min \left\{ \max \left\{ \mathbb{E}_P[Q(x,\xi)] \,:\, P \in \mathcal{P} \right\} \,:\, x \in \mathbb{X} \right\}.$$

Find the pair of optimal decision $x^* \in X$ which is good for all models $P \in \mathcal{P}$, among which there is a worst case model $P^* \in \mathcal{P}$.

The pair (x^*, P^*) forms a saddle point



Wasserstein distance

In order to measure the distance of two scenario distributions we use the transportation distance (Kantorovich distance, Wasserstein distance, earth mover distance) between random distributions on $\mathbb{R}^m = (\Omega, \mathrm{d})$ where d is a distance on $\mathbb{R}^m$.
Wasserstein distance of order r:

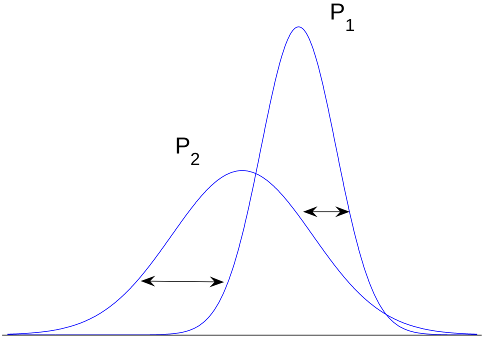
$$\mathsf{d}_r(\mathbb{P}_1, \mathbb{P}_2; d) := \left(\inf_\pi \left\{ \int_{\Omega\times\Omega} d(\omega_1, \omega_2)^r \pi[\mathrm{d}\omega_1, \mathrm{d}\omega_2] \right\} \right)^{\frac{1}{r}},$$

where the infimum is taken over all (bivariate) probability measures π on $\Omega \times \Omega$ which have respective marginals, thats

$$\pi[A \times \Omega] = \mathbb{P}_1[A] \text{ and } \pi[\Omega \times B] = \mathbb{P}_2[B]$$

for all measurable sets $A \subseteq \Omega$ and $B \subseteq \Omega$.
We shall call such a measure π a *transportation plan*.

Illustration of the Wasserstein distance



Multistage models

There is a multistage genberalization of the Wasserstein distance called nested distance (Pflug and Pichler, 2007).
As before, a baseline problem

$$\min \left\{ \mathcal{R}_{\hat{\mathbb{P}}}[Q(x,\xi)] \colon x \in \mathbb{X},\ x \lhd \mathfrak{F};\ \mathbb{P} = (\mathfrak{F}, P, \xi) \right\}$$

where the probability model is given by the nested distribution $\mathbb{P}$ for the stochastic process $\xi = (\xi_1, \ldots, \xi_T)$ is extended to the ambiguous model

$$\min_x \max_{\mathbb{P}} \left\{ \mathcal{R}_{\mathbb{P}}[Q(x,\xi)] \ :\ x \in \mathbb{X},\ x \lhd \hat{\mathfrak{F}};\ \mathbb{P} = (\hat{\mathfrak{F}}, P, \xi);\ \mathrm{dl}_r(\hat{\mathbb{P}}, \mathbb{P}) \leq \varepsilon \right\}.$$

The nested distance

The nested distance $\mathbb{d}_r(\mathbb{P},\bar{\mathbb{P}})$ is the minimal value of the following optimization program

$$\mathbb{d}_r(\mathbb{P},\bar{\mathbb{P}}) = \min\left\{\left[\int d_r^r(\xi,\bar{\xi})\,\pi(d\xi,d\bar{\xi})\right]^{1/r} : \ \pi \text{ fulfills } (\textbf{??}) \text{ and } (\textbf{??})\right\}$$

$$\pi(M\times\bar{\Omega}|\mathcal{F}_t\otimes\bar{\mathcal{F}}_t) = P(M|\mathcal{F}_t) \qquad M\in\mathcal{F}_T \tag{11}$$

$$\pi(\Omega\times N|\mathcal{F}_t\otimes\bar{\mathcal{F}}_t) = \bar{P}(N|\bar{\mathcal{F}}_t) \qquad N\in\bar{\mathcal{F}}_T. \tag{12}$$

The acceptable price for a contingent claim

Let C_t be a sequence of claims and let S_t be a sequence of hedging instruments. The acceptable ask-price for (C_t) is given as the optimal solution of the optimization problem

$$\begin{aligned} \pi_a(\mathcal{A}_1, \ldots, \mathcal{A}_T) := \min_{x,w} \; & w \\ \text{s.t. } & x_0^\top S_0 \leq w \\ & \mathcal{A}_t(x_t^\top S_t - x_{t-1}^\top S_t - C_t) \geq 0 \quad \forall t = 1, \cdots, T-1; \\ & \mathcal{A}_T(x_{T-1}^\top S_T - C_T) \geq 0. \end{aligned} \tag{13}$$

Characterization by dualization

Let the acceptability functionals $\mathcal{A}_t$ be positively homogeneous with supergradient set $\mathcal{Z}_t$, i.e.

$$\mathcal{A}_t(Y) = \inf\{\mathbb{E}[Y \cdot Z] : Z \in \mathcal{Z}_t\}.$$

Let $\tilde{S}_t$ be the discounted asset process and $\tilde{C}_t$ be the discounted payoff process. Let further

$$\mathcal{Q}^{\mathcal{A}} := \left\{ \mathbb{Q} : \mathbb{Q} \sim \mathbb{P};\ \mathbb{E}^{\mathbb{Q}}[\tilde{S}_{t+1}|\mathcal{F}_t] = \tilde{S}_t\ \forall t = 0, \dots T-1;\ \left.\frac{d\mathbb{Q}}{d\mathbb{P}}\right|_{\mathcal{F}_t} \in \mathcal{Z}_t \right\}.$$

Then

$$\pi_a(\mathcal{A}_1, \dots, \mathcal{A}_T) = \max\left\{ \sum_{t=1}^{T} \mathbb{E}^{\mathbb{Q}}\left[\tilde{C}_t\right] :\ \mathbb{Q} \in \mathcal{Q}^{\mathcal{A}} \right\}.$$

The ask price under ambiguity

The distributionally robust acceptable ask-price is defined as the optimal solution of the optimization problem

$$
\begin{aligned}
\min_{x_t, w} \quad & w \\
\text{s.t.} \quad & \\
& x_0^\top S_0 \le w \\
& \mathcal{A}_t^{\mathbb{P}}(x_{t-1}^\top S_t - x_t^\top S_t - C_t) \ge 0 \quad \forall \mathbb{P} \in \mathcal{P};\ \forall t = 1, \dots, T-1 \\
& \mathcal{A}_T^{\mathbb{P}}(x_{T-1}^\top S_T - C_T) \ge 0 \quad \forall \mathbb{P} \in \mathcal{P}
\end{aligned}
\tag{14}
$$

Dualization

Let $\mathcal{P}$ be finite set of models $\mathcal{P} = \{\mathbb{P}_1, \ldots, \mathbb{P}_n\}$. Then, the superreplication problem is by duality equivalent to

$$\sup_{\mathbb{Q}} \mathbb{E}^{\mathbb{Q}} \left[\sum_{t=1}^{T} \tilde{C}_t \right]$$

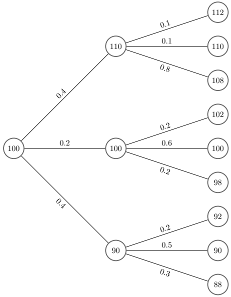
s.t.

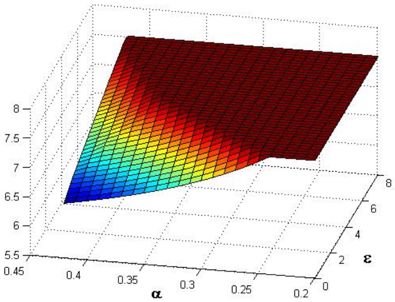
$$\mathbb{E}^{\mathbb{Q}} \left[\tilde{S}_{t+1} | \mathcal{F}_t \right] = \tilde{S}_t, \quad \forall t = 0, \ldots, T-1$$

$$\frac{d\mathbb{Q}}{d\hat{\mathbb{P}}}\Big|_{\mathcal{F}_t} \in \text{conv} \left\{ Z_t^{i,j} f_t^j \right\}, \quad \forall t = 0, \ldots, T$$

where $\hat{\mathbb{P}}$ is any model such that all $\mathbb{P}_j, j = 1, \ldots, n$, are absolutely continuous with $\frac{d\mathbb{P}_j}{d\hat{\mathbb{P}}} = f_j$, and $Z_t^i, i = 1, \ldots, k_t$, form the supergradient set of $\mathcal{A}_t$.

llustration for the ask-price





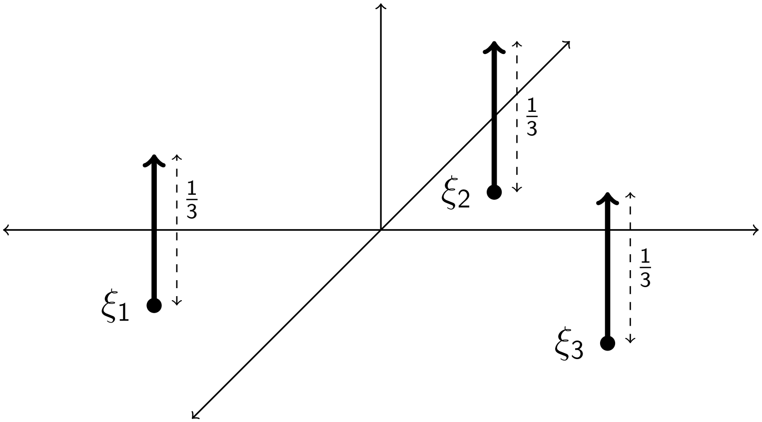
The ask price of a call option struck at 95 (on the ternary tree) as a function of the acceptance level α and the ambiguity radius ϵ.

Portfolio selection under model ambiguity

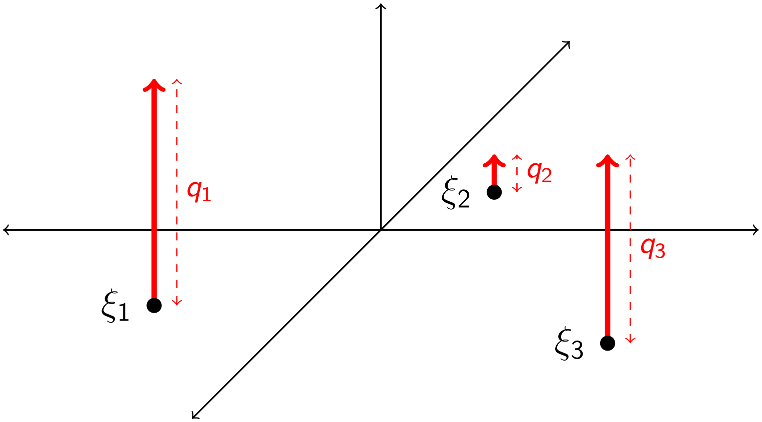
$$\max_{x\in\mathbb{X}} \min_{Q\in\mathbb{B}_\kappa(\hat{P})} \mathbb{E}\left(x^\top \xi^Q\right) - \lambda \mathrm{AV@R}_\alpha\left(-x^\top \xi^Q\right),$$

$\mathbb{B}_\kappa(P_0) := \{Q \in \mathcal{P}(\mathbb{R}^m) : \mathrm{d\!l}_1(Q, P_0) \le \kappa\}$,
$\hat{P}$... reference/baseline distribution,
κ... level of model ambiguity.
$\mathrm{d\!l}_p(\cdot,\cdot)$... Wasserstein distance of order p,
$\mathcal{P}(\mathbb{R}^m)$... space of all Borel probability measures on $\mathbb{R}^m$.

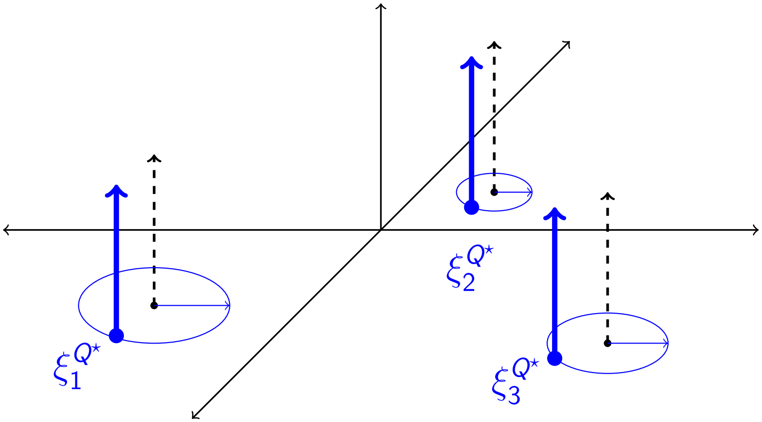
Consider the empirical distribution $P_0 = \hat{P}_n$



Pflug and Wozabal (2007)



Esfahani and Kuhn (2015)



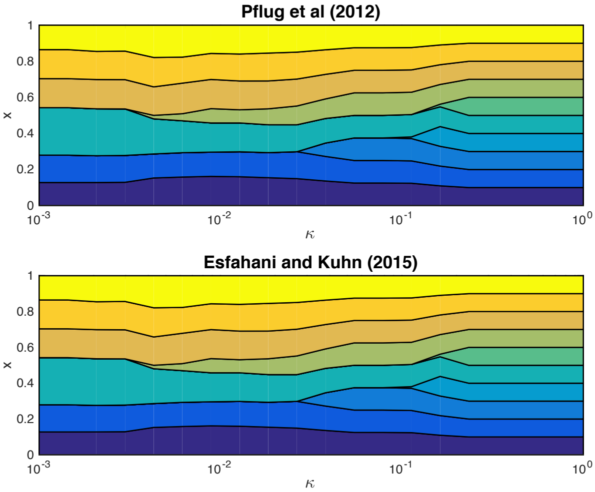


Figure: Optimal portfolio composition as a function of the level of model ambiguity κ.

Two different approaches

1. **Ambiguity in the *joint* distribution**
   $\Rightarrow$ Portfolio diversification
2. **Ambiguity in the dependence structure** with known marginal distribution
   $\Rightarrow$ Portfolio concentration

Total dependency ambiguity: Portfolio Concentration

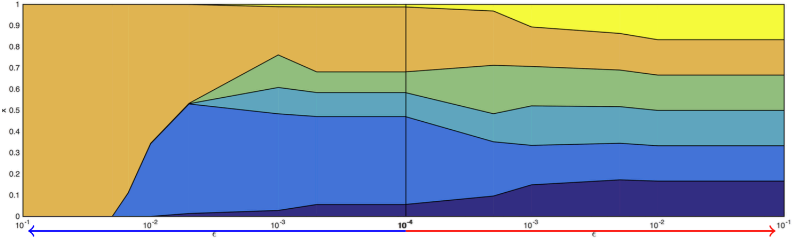
If $\mathcal{R}$ is (1) subadditive, (2) comonotone additive and (3) positive homogeneous, then

$$\max_{x\in\mathbb{X}} \min_{C\in\mathcal{C}} \mathbb{E}\left(-x^\top \xi^C\right) - \lambda\mathcal{R}\left(x^\top \xi^C\right)$$
$$= \max_{i\in\{1,\ldots,m\}} \mathbb{E}[\xi_i] - \lambda\mathcal{R}(\xi_i).$$

Thus the maximin portfolio is to invest everything in just one the asset i^*, where

$$i^* = \operatorname{argmax}_{i\in\{1,\ldots,m\}} \mathbb{E}[\xi_i] - \lambda\mathcal{R}(\xi_i).$$

Concentration vs Diversification



Ambiguity in the dependence structure Ambiguity in the joint distribution

Data: 6 Indices: S&P 500, TOPIX, FTSE China B35, EURO STOXX 50, FTSE 100 and NIFTY 500; observations Jan 1 - Dec 13, 2016

Equal weights is maximin for large ambiguity

With this insight, we may prove a remarkable result for distortion functionals:

$$\lim_{K\to\infty} \operatorname{argmax}_{\{\sum x_i=1, x_i\geq 0\}} \min_{\mathrm{d}_r(P,\hat{P})\leq K} \mathcal{U}_P(Y_x) = \frac{1}{M}\mathbb{1}.$$

Under large ambiguity, the optimal decision is the "equal weights" allocation.

The same result holds for the Markovitz model, if the distance is d_2.

Distortion utility functional: $\mathcal{U}(Y) = \int_0^1 F_Y(p)h(p)\,dp$

Average value-at-risk: $\mathbb{A}\mathrm{V@R}(Y) = \frac{1}{\alpha}\int_0^\alpha F_Y(p)\,dp$

Price of Ambiguity and Reward for Robustness

Let $\hat{\mathbb{P}}$ be the baseline model and let $x^*(\mathbb{P})$ be the optimal solution of the baseline problem. Likewise, let $\mathcal{P}$ be the ambiguity set and let $x^*(\mathcal{P})$ be the solution of the minimax problem. Under convex-concavity, the solution $x^*(\mathcal{P})$ of the minimax problem together with the worst case model $\mathbb{P}^*$ form a saddle point, meaning that the following inequality is valid for all feasible x and all $\mathbb{P} \in \mathcal{P}$

$$\mathbb{E}_{\mathbb{P}}[Q(x^*(\mathcal{P}), \xi)] \leq \mathbb{E}_{\mathbb{P}^*}[Q(x^*(\mathcal{P}), \xi)] \leq \mathbb{E}_{\mathbb{P}^*}[Q(x, \xi)].$$

Let us call $\mathbb{E}_{\mathbb{P}^*}[Q(x^*(\mathcal{P}), \xi)]$ the minimax value.

Define:

- The Price of Ambiguity.

$$\mathbb{E}_{\hat{\mathbb{P}}}[Q(x^*(\mathcal{P}),\xi)] - \mathbb{E}_{\hat{\mathbb{P}}}[Q(x^*(\hat{\mathbb{P}}),\xi)] \geq 0.$$

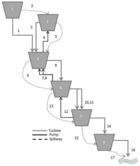
  "How much do I loose by implementing the minimax strategy $x^*(\mathcal{P})$ instead of the best strategy for the baseline model, if in fact the baseline model is true?"

- Reward for robust decisions.

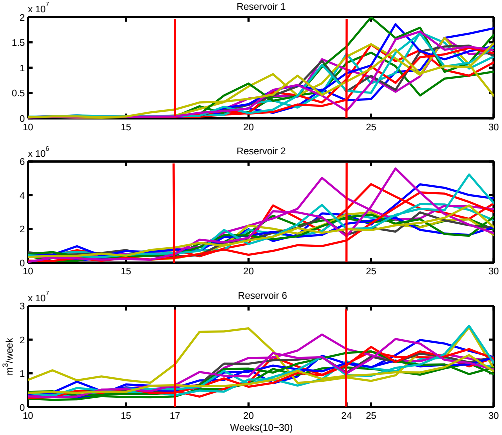
$$\mathbb{E}_{\mathbb{P}^*}[Q(x^*(\mathbb{P}),\xi)] - \mathbb{E}_{\mathbb{P}^*}[Q(x^*(\mathcal{P}),\xi)] \geq 0.$$

  "How much do I gain, when I implement the minimax strategy $x^*(\mathcal{P})$ instead of the best strategy for the baseline model, if in fact the worst case model is true?"

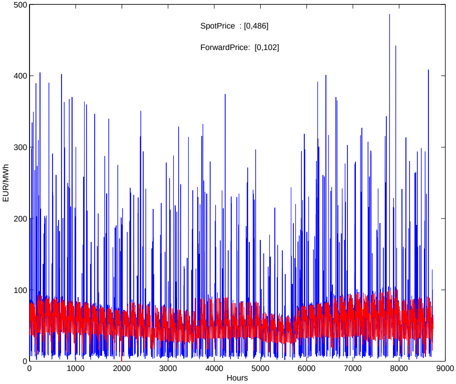
Management of a hydrosystem in the Austrian Alps



The scenario process consist of 5 components: Spot prices, Pumping prices, Inflows for 3 reservoirs. Statistical model selection methods were used to find that the inflows can be represented by a 3-dimensional $SARMA(1,2),(2,2)_5 2$ process, while the spot and pumping prices can be modeled by an independent process, a superposition of an additive error model based on forward prices and a spike generating process.



Observations for Inflows


SpotPrice : [0,486]
ForwardPrice: [0,102]
EUR/MWh
Hours
0
100
200
300
400
500
0
1000
2000
3000
4000
5000
6000
7000
8000
9000

The decision model

maximize

$\lambda\,\mathbb{E}[x_T^c] - (1-\lambda)\mathbb{A}\mathrm{V@R}_{1-\alpha}[-x_T^c]$

subject to

$0 \leq x_{t,i}^f \leq \overline{x}_i^f,$

$\underline{x}_j^s \leq x_{t,j}^s \leq \overline{x}_j^s,$

$x_{end,j}^{,s} \leq x_{T,j}^s,$

$x_{t,j}^s = x_{t-1,j}^s + \xi_{t,j}^f + \sum_{\{i \in I | P_{max} > 0\}} A_{i,j} \cdot x_{t-1,i}^f + \sum_{\{i \in I | P_{max} = 0\}} A_{i,j} \cdot x_{t,i}^f,$

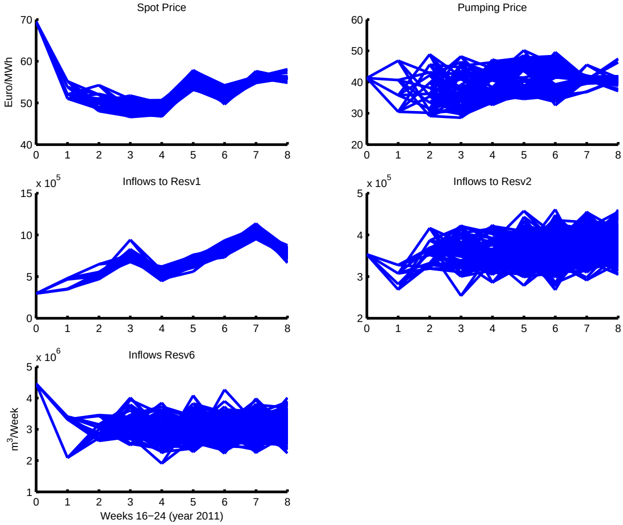
$x_{t,i}^e = x_{t-1,i}^f \cdot k^i \cdot \triangle t_{(t-1)},$

$x_t^c = x_{t-1}^c \cdot (1+r)^{\triangle t_{(t-1)}} + \sum_{\{i \in I | k^i > 0\}} x_{t-1,i}^e \cdot \xi_t^e + \sum_{\{i \in I | k^i < 0\}} x_{t-1,i}^i \cdot \xi_t^p.$

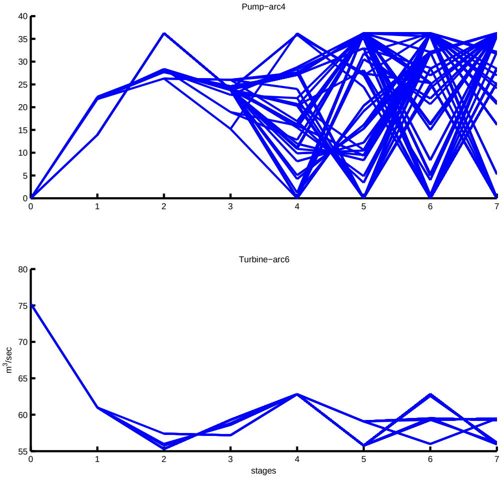
Generating a scenario tree

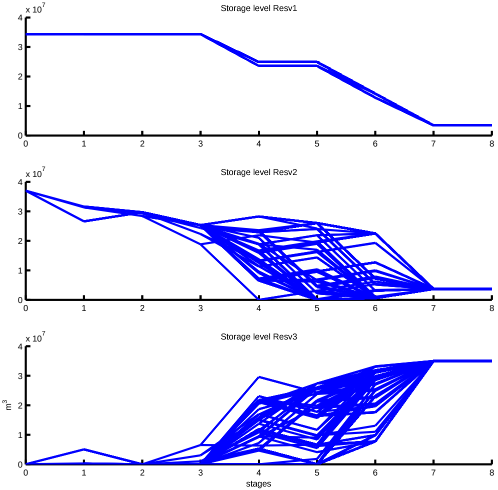
We generate a scenario tree in a way that the nested distance between the scenario process and the scenario tree is as small as possible.

Number of stages	8
Minimal bushiness per stage	2,2,2,1,1,1,1,1
Maximal distance per stage	5,5,5,7,7,7,10,10
Number of scenarios (leaves)	392
Number of nodes 1532	

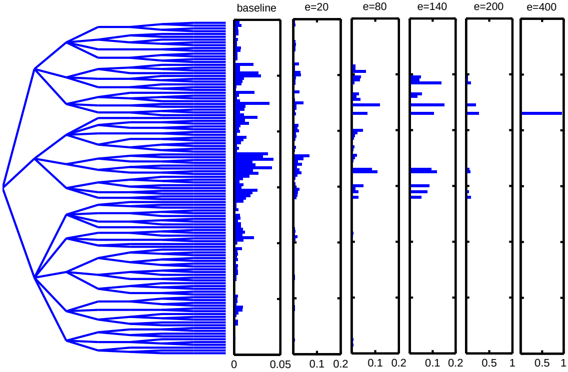


Spot Price
Euro/MWh
Pumping Price
Inflows to Resv1
x 10^5
Inflows to Resv2
x 10^5
Inflows Resv6
x 10^6
m^3/Week
Weeks 16−24 (year 2011)

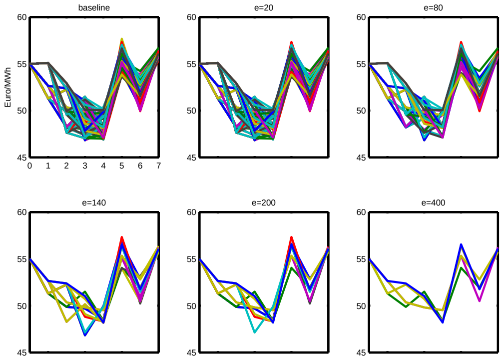

Pump−arc4
40
35
30
25
20
15
10
5
0
0 1 2 3 4 5 6 7
Turbine−arc6
80
75
70
65
60
55
m3/sec
0 1 2 3 4 5 6 7
stages



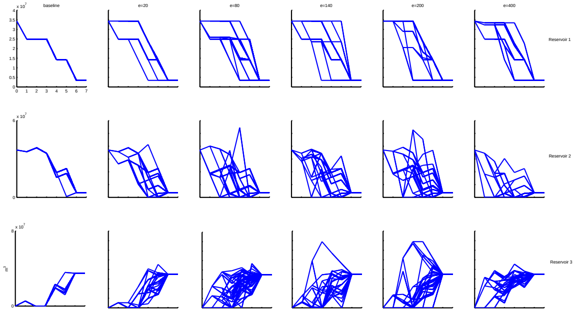
Storage level Resv1
x 10^7
Storage level Resv2
x 10^7
Storage level Resv3
x 10^7
m^3
stages



The typical picture: The larger is the ambiguity radius, the simpler

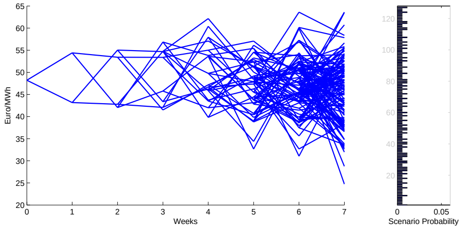


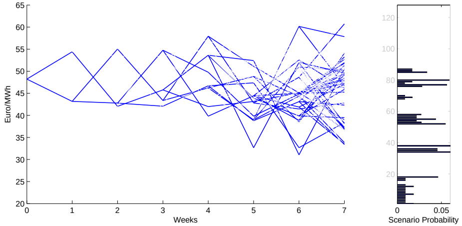
The worst case spotprice trees: Only the arcs with a minimum



The minimax decisions: They get more complicated with increasing ambiguity radius: Decisions lying on bounds are avoided.
Price of ambiguity: 2.3%.
Reward for robustness: 7.5%.

Worst case tree for a thermal plant optimization





Conclusions

- In order to capture scenario uncertainty (aleatoric uncertainty) and probability ambiguity (epistemic uncertainty-model error) we use a probabilistic maximin approach.
- The ambiguity neighborhood should be chosen in such a way that it corresponds to statistical confidence regions for which bounds for the covering probability are available.
- If the ambiguity radius is increased, then the saddle point changes typically in the following way:
  - The robust decision strategy becomes more complicated and "diversified"
  - The worst case model gets more simpler
- It turns out that the price to be paid for including ambiguity in the optimization problem is often smaller than the reward one gets for robustifying the solution.