Untying the Knot between a Stochastic Program and its Distribution

Erick Delage
Assistant professor
Service des Méthodes Quantitatives de Gestion
HEC Montréal

Tuesday, August 16th, 2011

Evidence that Managing an Investment Portfolio is Difficult

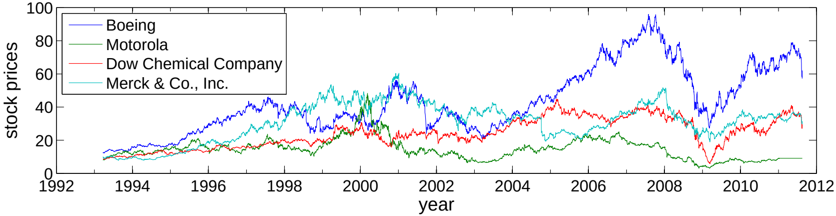
Value on Jan 1st 2009 of each dollar contribution made to the Caisse de Dépôts et de Placements

Date of contribution	CDPQ	1-year guaranteed certificates
Jan 1st, 2008	\$0,75	\$1,03
Jan 1st, 2007	\$0,79	\$1,05
Jan 1st, 2006	\$0,91	\$1,07
Jan 1st, 2005	\$1,04	\$1,09
Jan 1st, 2004	\$1,17	\$1,10
Jan 1st, 2003	\$1,35	\$1,12
Jan 1st, 2002	\$1,22	\$1,13
Jan 1st, 2001	\$1,16	\$1,18
Jan 1st, 2000	\$1,23	\$1,23

Why are Financial Investments so Fragile?

Some reasons:

- A wide range of financial securities can be used for investment
- Securities have become very complex
- The risks involved are difficult to evaluate
- Limited knowledge of how the market will behave in the future



Are Airlines Adventurous in their Fleet Acquisition?

- Fleet composition is a difficult decision problem:
  - Fleet contracts are signed 10 to 20 years ahead of schedule.
  - Many factors are still unknown at that time:
    e.g., passenger demand, fuel prices, etc.
- Yet, most airline companies sign these contracts based on a single scenario of what the future may be.
- Are airlines companies at risk of going bankrupt?

Stochastic Programming Approach

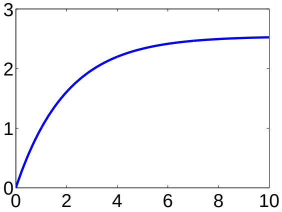
Let's consider the stochastic programming problem:

$$\underset{\mathbf{x}\in\mathcal{X}}{\text{maximize}} \quad \mathbb{E}\left[u(\mathrm{h}(\mathbf{x},\boldsymbol{\xi}))\right]$$

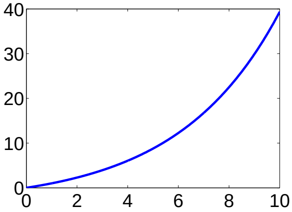
where $\mathbf{x}$ = decisions and $\boldsymbol{\xi}$ = uncertain parameters.

Here, we assume that we know:

- The distribution of the random vector $\boldsymbol{\xi}$
- A utility function that matches investor's attitude to risk



Retirement Fund



Vegas Budget

Difficulty of developing a probabilistic model

Developing an accurate probabilistic model requires heavy engineering efforts:

- Need to collect enough observations
- Need to consult with experts of the field of practice
- Need to make simplifying assumptions

Yet, there are inherent pitfalls in the process:

- Expecting that a scenario might occur does not determine its probability of occurring
- Unexpected event (e.g., economic crisis) might occur
- The future might actually not behave like the past

Limits of Expected Utility: Ellsberg Paradox

Consider an urn with 30 blue balls and 60 other balls that are either red or yellow (you don't know how many are red or yellow).

Experiment 1: Choose among the following two gambles

- Gamble A: If you draw a blue ball, then you win 100$
- Gamble B: If you draw a red ball, then you win 100$

Experiment 2: Choose among the following two gambles

- Gamble C: If you draw blue or yellow ball, then you win 100$
- Gamble D: If you draw red or yellow ball, then you win 100$

If you clearly prefer Gamble A & D, then you cannot be thinking in terms of expected utility.

Untying the SP from a Specific Distribution

- Let's consider that the choice of F is ambiguous
- Use available information to define $\mathcal{D}$, such that $F \in \mathcal{D}$
- We are faced with a multi-objective optimization problem:

$$\underset{\mathbf{x}\in\mathcal{X}}{\text{maximize}} \quad \{\, \mathbb{E}_F[u(\mathsf{h}(\mathbf{x},\boldsymbol{\xi}))] \,\}_{F\in\mathcal{D}}$$

- Distributionally Robust Optimization values the lowest performing one

$$(DRSP) \qquad \underset{\mathbf{x}\in\mathcal{X}}{\text{maximize}} \; \min_{F\in\mathcal{D}} \mathbb{E}_F[u(\mathsf{h}(\mathbf{x},\boldsymbol{\xi}))]$$

  - Introduced by H. Scarf in 1958
  - Recently, we found ways of solving some DRSP's efficiently [Popescu (2007), Bertsimas et al., Natarajan et al., Delage et al. (2010)]
  - Possible to promote performance differently depending on F [Föllmer et al. (2002), Li et al. (2011)]

Outline

1. Introduction
2. Distributionally Robust Optimization
3. Distributions Can Be Misleading
4. Value of Stochastic Modeling
5. Conclusion

Outline

1. Introduction
2. Distributionally Robust Optimization
3. Distributions Can Be Misleading
4. Value of Stochastic Modeling
5. Conclusion

Assumptions on Objective Function

Let's make two assumptions about $\mathbb{E}[u(\mathrm{h}(\mathbf{x},\boldsymbol{\xi}))]$.

1. The utility function is piecewise linear concave :

$$u(y) = \min_{1\leq k\leq K} \quad a_k y + b_k \quad ,$$

2. The profit function is the maximum of a linear program with uncertainty limited to objective

$$\begin{aligned} \mathrm{h}(\mathbf{x},\boldsymbol{\xi}) \;:=\; \max_{\mathbf{y}}. \quad & \mathbf{c}_1^{\mathsf{T}}\mathbf{x} + \boldsymbol{\xi}^{\mathsf{T}} C_2 \mathbf{y} \\ \text{s.t.} \quad & A\mathbf{x} + B\mathbf{y} \leq \mathbf{b} \end{aligned}$$

Resolving Distributional Set from Data

- Question:
  - We have in hand i.i.d. samples $\{\boldsymbol{\xi}_i\}_{i=1}^M$
  - We know that $\mathbb{P}(\boldsymbol{\xi} \in \mathcal{S}) = 1$ and $\mathcal{S} \subseteq \mathcal{B}(\mathbf{0}, R)$
  - We can estimate the mean and covariance matrix:

$$\hat{\boldsymbol{\mu}} = \frac{1}{M}\sum_{i=1}^M \boldsymbol{\xi}_i \qquad \hat{\Sigma} = \frac{1}{M}\sum_{i=1}^M (\boldsymbol{\xi}_i - \hat{\boldsymbol{\mu}})(\boldsymbol{\xi}_i - \hat{\boldsymbol{\mu}})^\mathsf{T}$$

  - What do we know about the distribution behind these samples?

- Answer:

$$\mathcal{D}(\gamma) = \left\{ F \left| \begin{array}{l} \mathbb{P}(\boldsymbol{\xi} \in \mathcal{S}) = 1 \\ \|\mathbb{E}\left[\boldsymbol{\xi}\right] - \hat{\mu}\|^2_{\hat{\Sigma}^{-1/2}} \leq \gamma_1 \\ \mathbb{E}\left[(\boldsymbol{\xi} - \hat{\boldsymbol{\mu}})(\boldsymbol{\xi} - \hat{\boldsymbol{\mu}})^\mathsf{T}\right] \preceq (1+\gamma_2)\hat{\Sigma} \end{array} \right. \right\}$$

  - With prob. $> 1 - \delta$ the distribution is contained in $\mathcal{D}(\gamma)$ for some $\gamma_1 = O\left(\frac{R^2}{M}\log(1/\delta)\right)$ and $\gamma_2 = O\left(\frac{R^2}{\sqrt{M}}\sqrt{\log(1/\delta)}\right)$.

The DRSP is a SDP

- The DRSP problem with $\mathcal{D}(\gamma)$ is equivalent to

$$\begin{array}{lll} \max\limits_{\mathbf{x},\,\mathbf{Q},\mathbf{q},r} & r - \left(\gamma_2\hat{\Sigma} + \hat{\boldsymbol{\mu}}\hat{\boldsymbol{\mu}}^{\mathsf{T}}\right) \bullet \mathbf{Q} - \hat{\boldsymbol{\mu}}^{\mathsf{T}}\mathbf{q} - \sqrt{\gamma_1}\,\|\hat{\Sigma}^{1/2}(\mathbf{q} + 2\mathbf{Q}\hat{\boldsymbol{\mu}})\| \\ \text{s.t.} & r \leq \min\limits_{\boldsymbol{\xi}\in\mathcal{S}} u(\mathrm{h}(\mathbf{x},\boldsymbol{\xi})) + \boldsymbol{\xi}^{\mathsf{T}}\mathbf{q} + \boldsymbol{\xi}^{\mathsf{T}}\mathbf{Q}\boldsymbol{\xi} \quad (\star) \\ & \mathbf{Q} \succeq 0 \end{array}$$

- If $\mathcal{S}$ = polygon or ellipsoid, then DRSP equivalent to semi-definite program.
  E.g., when $\mathcal{S} = \mathbb{R}^m$, Constraint $(\star)$ can be replaced by

$$\begin{bmatrix} \mathbf{Q} & (\mathbf{q} + a_k C_2 y_k)/2 \\ (\mathbf{q} + a_k C_2 y_k)^{\mathsf{T}}/2 & a_k c_1^{\mathsf{T}}\mathbf{x} + b_k - r \end{bmatrix} \succeq 0\ ,\ \forall k$$

The Robustness of the Deterministic Solution

If we are risk neutral we might not even need distribution information

Theorem

The solution of

$$\underset{\mathbf{x}\in\mathcal{X}}{\text{maximize}} \quad \mathbb{E}[h(\mathbf{x},\mu)]$$

is optimal with respect to

$$\underset{\mathbf{x}\in\mathcal{X}}{\text{maximize}} \quad \inf_{F\in\mathcal{D}(\mu,\Psi)} \quad \mathbb{E}_F[h(\mathbf{x},\xi)] \ ,$$

for any set of convex functions Ψ *with*

$$\mathcal{D}(\mu,\Psi) = \left\{ F \,\middle|\, \begin{array}{l} \mathbb{E}[\xi] = \mu \\ \mathbb{E}[\psi(\xi)] \leq 0 \ , \ \forall\, \psi \in \Psi \end{array} \right\} .$$

Outline

1. Introduction
2. Distributionally Robust Optimization
3. Distributions Can Be Misleading
4. Value of Stochastic Modeling
5. Conclusion

Distributionally Robust Portfolio Optimization

Let's consider the case of portfolio optimization:

$$\max_{\mathbf{x}\in\mathcal{X}} . \quad \min_{F\in\mathcal{D}} \mathbb{E}_F[u(\boldsymbol{\xi}^\mathsf{T}\mathbf{x})] \quad ,$$

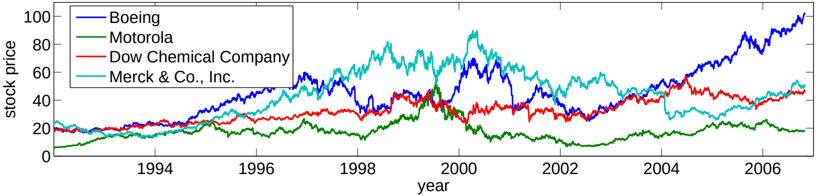
where x_i is how much is invested in stock i with future return ξ_i.

Does the robust solution perform better than a stochastic programming solution?

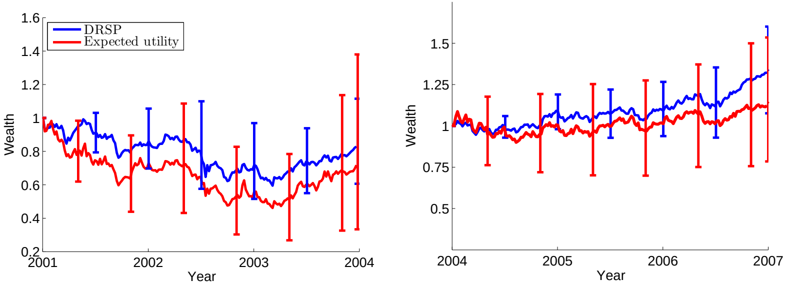
$$\mathcal{D} = \mathcal{D}(\gamma) \qquad \text{vs.} \qquad \mathcal{D} = \{\hat{F}\}$$

Experiments in Portfolio Optimization

30 stocks tracked over years 1992-2007 using Yahoo! Finance



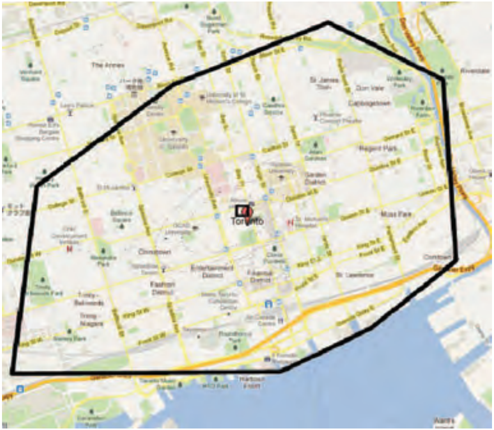
Wealth Evolution for 300 Experiments



- 10% and 90% percentiles are indicated periodically.
- 79% of time, the DRSP outperformed the exp. utility model
- 67% improvement on average using DRSP with $\mathcal{D}(\gamma)$

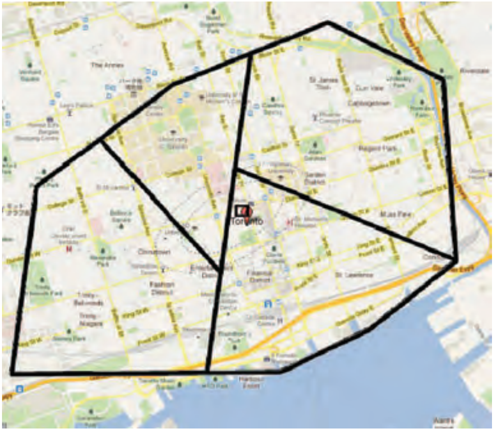
Multi-Vehicle Routing on a Planar Region

- Divide a planar region into K subregions, each serviced by a different vehicle, so that the total workload be most evenly distributed among the fleet



Multi-Vehicle Routing on a Planar Region

- Divide a planar region into K subregions, each serviced by a different vehicle, so that the total workload be most evenly distributed among the fleet



Multi-Vehicle Routing on a Planar Region

- Divide a planar region into K subregions, each serviced by a different vehicle, so that the total workload be most evenly distributed among the fleet

Multi-Vehicle Routing on a Planar Region

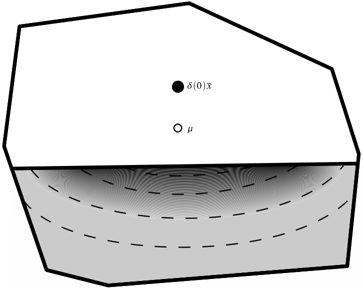
- Divide a planar region into K subregions, each serviced by a different vehicle, so that the total workload be most evenly distributed among the fleet

Distributionally Robust Partitioning

- Given $\mathcal{D}$, we partition so that the largest workload over the worst distribution of demand points is as small as possible

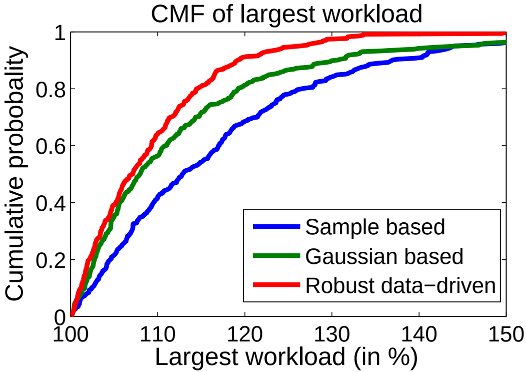
$$\min_{\{\mathcal{R}_1,\mathcal{R}_2,\ldots,\mathcal{R}_K\}} \sup_{F\in\mathcal{D}} \left\{ \max_i \mathbb{E}[TSP(\{\xi_1,\xi_2,\ldots,\xi_N\}\cap\mathcal{R}_i)] \right\} \quad ,$$

- A side product is to characterize for any partition what is a worst-case distribution of demand locations



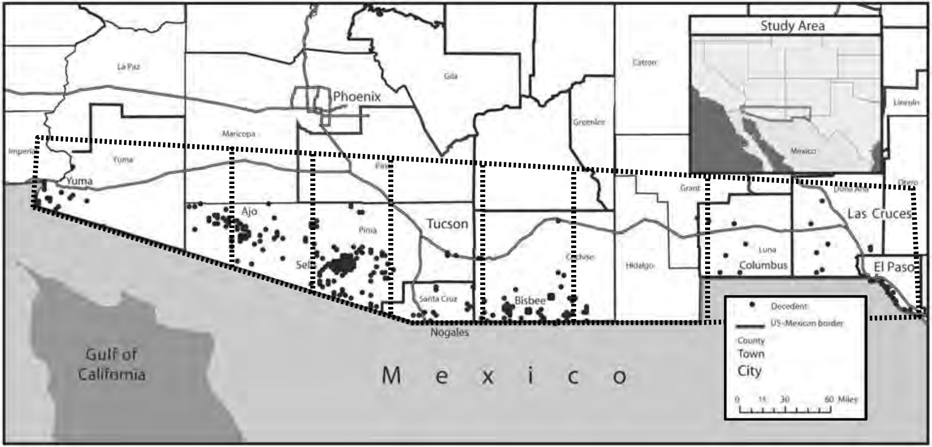
Distributionally Robust Partitioning

We simulated three partition schemes on a set of randomly generated parcel delivery problems where the territory needed to be divided into two regions and the demand is drawn from a mixture of truncated Gaussian distribution



Border Patrol Workload Partitioning

Robust partitions of the USA-Mexico border obtained using our branch & bound algorithm.



Outline

1. Introduction
2. Distributionally Robust Optimization
3. Distributions Can Be Misleading
4. **Value of Stochastic Modeling**
5. Conclusion

The Value of Stochastic Modeling

Consider the situation:

1. We know of a set $\mathcal{D}$ such that $F \in \mathcal{D}$
2. We have a candidate solution $\mathbf{x}_1$ in mind
3. Is it worth developing a stochastic model: $\mathcal{D} \rightarrow F$?
   - (a) If yes, then develop a model & solve it
   - (b) Otherwise, implement $\mathbf{x}_1$

The Value of Stochastic Modeling ($\mathcal{VSM}$) gives an optimistic estimate of the value of obtaining perfect information about F.

$$\mathcal{VSM}(\mathbf{x}_1) \;:=\; \sup_{F \in \mathcal{D}} \left\{ \max_{\mathbf{x}_2} \mathbb{E}_F[\mathrm{h}(\mathbf{x}_2, \boldsymbol{\xi})] - \mathbb{E}_F[\mathrm{h}(\mathbf{x}_1, \boldsymbol{\xi})] \right\}$$

Theorem

Unfortunately, evaluating $\mathcal{VSM}(\mathbf{x}_1)$ exactly is NP-hard in general.

Bounding the Value of Stochastic Modeling

Theorem

If $\mathcal{S} \subseteq \{\boldsymbol{\xi} \,|\, \|\boldsymbol{\xi}\|_1 \leq \rho\}$*, an upper bound can be evaluated in* $O(d^{3.5} + d\, T_{DCP})$ *using:*

$$\begin{aligned} \mathcal{UB}(\mathbf{x}_1, \bar{\mathbf{y}}_1) \;:=\; \min_{s,\mathbf{q}} \quad & s + \boldsymbol{\mu}^\mathsf{T}\mathbf{q} \\ \text{s.t.} \quad & s \geq \alpha(\rho\mathbf{e}_i) - \rho\mathbf{e}_i^\mathsf{T}\mathbf{q} \;, \forall\, i \in \{1, ..., d\} \\ & s \geq \alpha(-\rho\mathbf{e}_i) + \rho\mathbf{e}_i^\mathsf{T}\mathbf{q} \;, \forall\, i \in \{1, ..., d\}\,, \end{aligned}$$

where $\alpha(\boldsymbol{\xi}) = \max_{\mathbf{x}_2} h(\mathbf{x}_2, \boldsymbol{\xi}) - h(\mathbf{x}_1, \boldsymbol{\xi}; \bar{\mathbf{y}}_1)$.

- $\mathcal{UB}$ only uses information about $\boldsymbol{\mu}$ and $\mathcal{S}$
- $\mathcal{UB}$ simplifies the structure of $\mathcal{S}$
- $\mathcal{UB}$ assumes the candidate decision $\mathbf{y}_1$ cannot adapt to $\boldsymbol{\xi}$

Mathematical formulation for Fleet Mix Problem

The fleet composition problem is a stochastic mixed integer LP

Fleet mix

$$\underset{\mathbf{x}}{\max.}\quad \mathbb{E}\,[-\underbrace{\mathbf{o}^{\mathsf{T}}\mathbf{x}}_{\text{ownership cost}} + \underbrace{h(\mathbf{x},\tilde{\mathbf{p}},\tilde{\mathbf{c}},\tilde{\mathbf{L}})}_{\text{future profits}}\,]\ ,$$

with $h(\mathbf{x},\tilde{\mathbf{p}},\tilde{\mathbf{c}},\tilde{\mathbf{L}}) :=$

$$\max_{z\geq 0,y\geq 0,w} \quad \sum_k \Big(\sum_i \overbrace{\tilde{p}_i^k w_i^k}^{\text{flight profit}} - \overbrace{\tilde{c}_k(z_k - x_k)^+}^{\text{rental cost}} + \overbrace{\tilde{L}_k(x_k - z_k)^+}^{\text{lease revenue}} \Big)$$

$$\text{s.t.} \quad \left. w_i^k \in \{0,1\}\ ,\ \forall\, k, \forall\, i \quad \& \quad \sum_k w_i^k = 1\ ,\ \forall\, i \right\}\ \text{Cover}$$

$$\left. y^k_{g\in\text{in}(v)} + \sum_{i\in\text{arr}(v)} w_i^k = y^k_{g\in\text{out}(v)} + \sum_{i\in\text{dep}(v)} w_i^k\ ,\ \forall\, k, \forall\, v \right\}\ \text{Balance}$$

$$\left. z_k = \sum_{v\in\{v|\text{time}(v)=0\}} \big(y^k_{g\in\text{in}(v)} + \sum_{i\in\text{arr}(v)} w_i^k\big)\ ,\quad \forall k \right\}\ \text{Count}$$

Experiments in Fleet Mix Optimization

We experimented with three test cases :

1. 3 types of aircrafts, 84 flights, $\sigma_{\tilde{p}_i}/\mu_{\tilde{p}_i} \in [4\%, 53\%]$
2. 4 types of aircrafts, 240 flights, $\sigma_{\tilde{p}_i}/\mu_{\tilde{p}_i} \in [2\%, 20\%]$
3. 13 types of aircrafts, 535 flights, $\sigma_{\tilde{p}_i}/\mu_{\tilde{p}_i} \in [2\%, 58\%]$

Results:

Test cases	Computation Times			Upper bound for VSM
	DCP	SP (100 scen.)	$\mathcal{UB}$	
#1	0.6 s	3 min	12 sec	6%
#2	1 s	14 min	40 sec	1%
#3	5 s	21 h	2 min	7%

Conclusions:

- It's wasteful to invest more than 7% of profits in extra info

Outline

1. Introduction
2. Distributionally Robust Optimization
3. Distributions Can Be Misleading
4. Value of Stochastic Modeling
5. Conclusion

Conclusion & Future Work

- Many forms of the DRSP are tractable
- Some actually reduce to the DCP
- Thinking we know the distribution can be misleading
- Knowing the actual distribution might not help that much
- There are tools that help estimate how much the true distribution is worth

- Open questions :
  - Can tractable DRSP be made consistent ?
  - Can DRSP be extended to multi-objective problems ?
  - How to deal with ambiguity about one's utility function ?

Bibliography

Armbruster, B., E. Delage. 2011. Decision making under uncertainty when preference information is incomplete. Working paper.

Bertsimas, D., X. V. Doan, K. Natarajan, C. P. Teo. 2010. Models for minimax stochastic linear optimization problems with risk aversion. *Mathematics of Operations Research* **35**(3) 580–602.

Carlsson, J. G., E. Delage. 2011. Robust partitioning for stochastic multi-vehicle routing. Working paper.

Delage, E., S. Arroyo, Y. Ye. 2011. The value of stochastic modeling in two-stage stochastic programs with cost uncertainty. Working paper.

Delage, E., Y. Ye. 2010. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research* **58**(3) 595–612.

Natarajan, K., M. Sim, J. Uichanco. 2010. Tractable Robust Expected Utility and Risk Models for Portfolio Optimization. *Mathematical Finance* **20**(4) 695–731.

Questions & Comments ...

... Thank you!