

Bias, Exploitation and Proxies in Scenario-Based Risk Minimization

Helmut Mausser
(jointly with Oleksandr Romanko)

October 3, 2011

Algorithmics



Motivation

Consider the relative loss in value (“loss return”) of a portfolio in basis points

Minimize risk (variance, CVaR) as estimated from a random sample of 1,000 return scenarios and evaluate actual risk from 1,000,000 return scenarios

Min.	In-Sample	Out-of-Sample
Variance	4.37	4.42
CVaR (90%)	293.30	304.45
CVaR (95%)	344.60	379.41
CVaR (99%)	412.48	525.74
CVaR (99.9%)	417.01	712.89

Motivation

Consider the relative loss in value (“loss return”) of a portfolio in basis points

Minimize risk (variance, CVaR) as estimated from a random sample of 1,000 return scenarios and evaluate actual risk from 1,000,000 return scenarios

Eval. Min.	In-Sample (1,000 scenarios)					Actual (1,000,000 scenarios)				
	Variance	CVaR (90%)	CVaR (95%)	CVaR (99%)	CVaR (99.9%)	Variance	CVaR (90%)	CVaR (95%)	CVaR (99%)	CVaR (99.9%)
Variance	4.37	323.00	391.95	490.19	593.81	4.42	312.14	376.40	501.10	647.73
CVaR (90%)	5.00	293.30	357.59	459.86	501.96	5.08	304.45	373.56	508.00	663.12
CVaR (95%)	5.04	301.08	344.60	454.36	499.54	5.17	309.64	379.41	515.42	674.77
CVaR (99%)	5.52	311.34	365.69	412.48	493.71	5.37	315.63	386.74	525.74	685.92
CVaR (99.9%)	5.89	317.94	375.80	417.01	417.01	5.85	328.24	402.38	546.95	712.89

“Natural estimator” (diagonal) always best in-sample but not out-of-sample

Optimism

optimism = actual risk – perceived risk

Eval.		Optimism				
Opt.	Variance	CVaR (90%)	CVaR (95%)	CVaR (99%)	CVaR (99.9%)	
Variance	0.05	-10.86	-15.55	10.91	53.92	
CVaR (90%)	0.08	11.15	15.97	48.14	161.16	
CVaR (95%)	0.13	8.56	34.81	61.06	175.23	
CVaR (99%)	-0.15	4.29	21.05	113.26	192.21	
CVaR (99.9%)	-0.04	10.30	26.58	129.94	295.88	

A “proxy” estimator that is less optimistic than the natural estimator may perform better out-of-sample

Notation

$w \in \Omega \subseteq \mathbb{R}^J$ is a portfolio, where w_j is the weight of asset j

\mathfrak{F} is the multivariate distribution of asset returns

$\pi(w)$ is the actual risk of the portfolio w

$p_N(w, S)$ is an estimate of $\pi(w)$ based on a sample S of size N from \mathfrak{F}

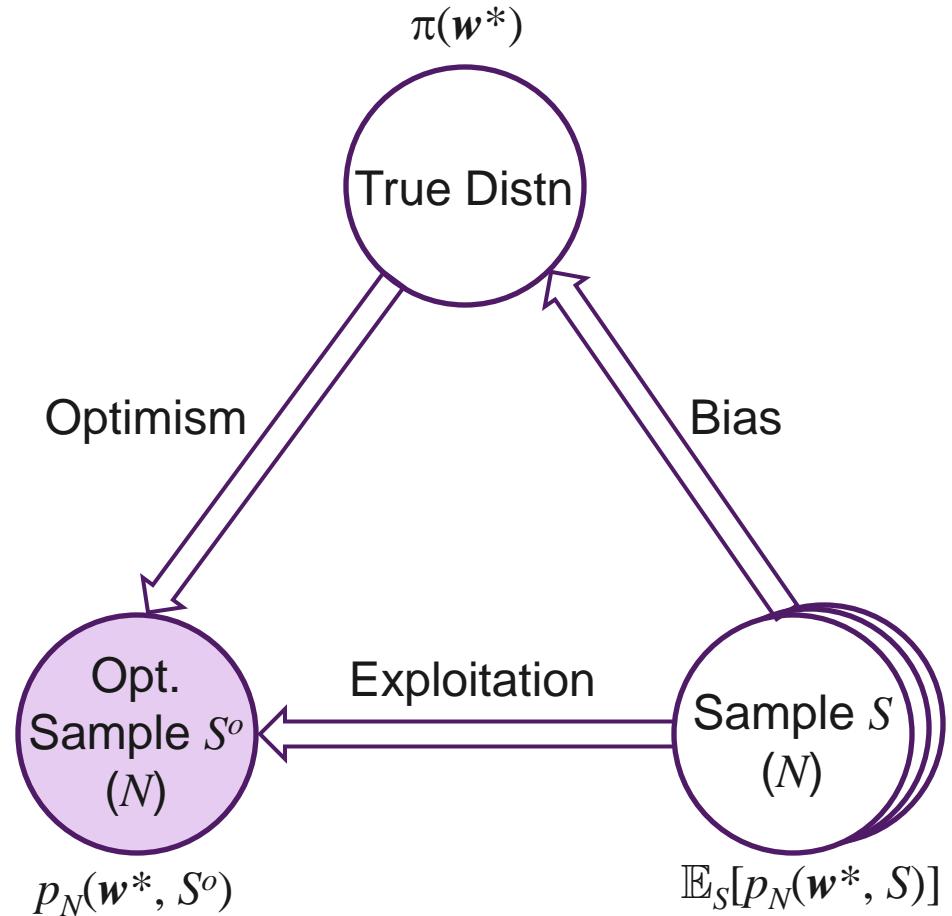
Goal is to obtain w^π such that $\pi(w^\pi)$ is minimal

- Generate a random sample S^o comprising N observations (scenarios) from \mathfrak{F}
- Find $w^* \equiv w^*(b_N, S^o)$ that minimizes $b_N(w, S^o)$ for some estimator b_N
 - $b_N = p_N$ is the natural estimator
 - $b_N \neq p_N$ is a proxy
- w^* has perceived risk $p_N(w^*, S^o)$ and actual risk $\pi(w^*)$

Exploitation and Bias

- Exploitation of S^o
 - Optimization takes advantage of sampling noise (e.g., “overfitting”)
 - Depends on amount of noise, flexibility of formulation, functional form of risk measure
- Bias of p_N
 - Sample variance, sample mean are unbiased but CVaR and VaR estimators are biased (for finite samples)
 - Depends on underlying distribution (i.e., w^*) and sample size

$$\text{optimism} = \text{exploitation} - \text{bias}$$

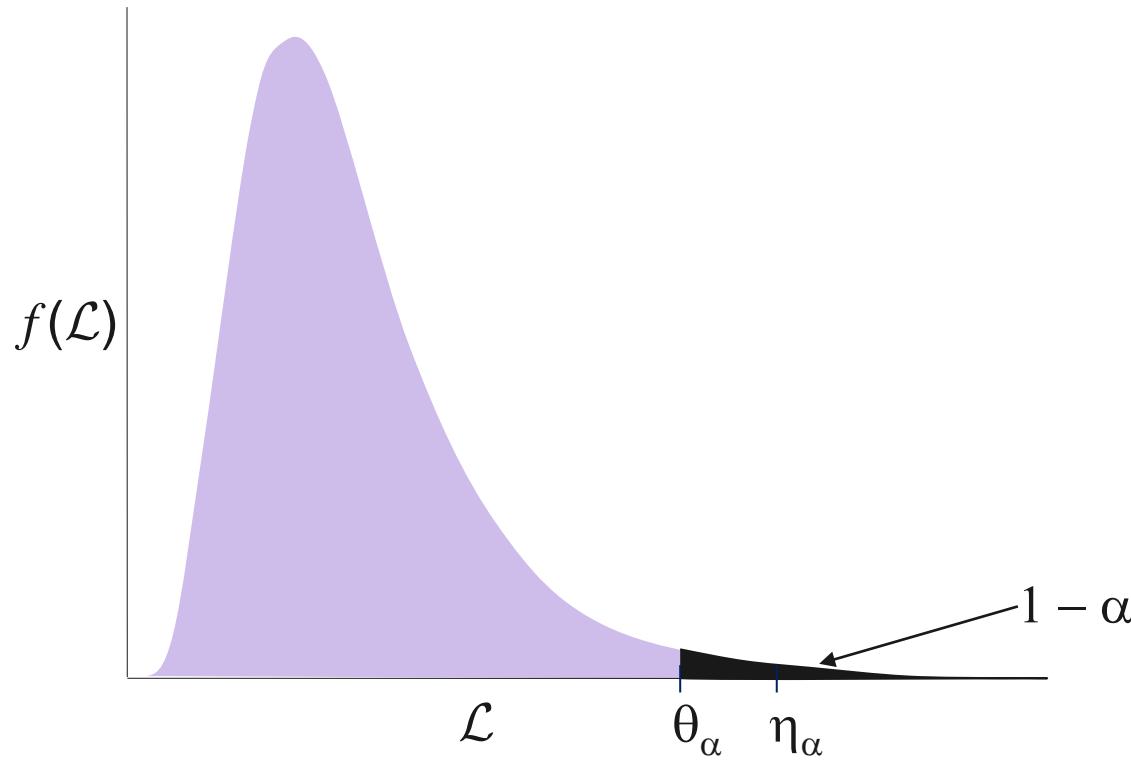


Lim, A.E.B., Shantikumar, J.G. and G.-Y. Vahn (2011), “Fragility of CVaR in Portfolio Optimization,” *Operations Research Letters* 39, 163-171.

Tail-Based Risk Measures

Consider a continuous random variable \mathcal{L} with distribution F

- The Value-at-Risk (VaR) at level α : $\theta_\alpha = F^{-1}(\alpha)$
- The Conditional Value-at-Risk (CVaR) at level α : $\eta_\alpha = \mathbb{E}[\mathcal{L} | \mathcal{L} > \theta_\alpha]$



Estimators

Given a random sample of size N , let $\ell_{(k)}$ be the k^{th} order statistic, i.e.,
 $\ell_{(1)} \leq \ell_{(2)} \leq \dots \leq \ell_{(N)}$

- An estimate of θ_α is $q_{\alpha,N} = \ell_{(\lceil N\alpha \rceil)}$

- An estimate of η_α is $h_{\alpha,N} = \frac{1}{N(1-\alpha)} \left[(\lceil N\alpha \rceil - N\alpha) \ell_{(\lceil N\alpha \rceil)} + \sum_{k=\lceil N\alpha \rceil+1}^N \ell_{(k)} \right]$

...	$\ell_{(98)}$	$\ell_{(99)}$	$\ell_{(100)}$	$q_{0.98,100} = 0.42$	$h_{0.98,100} = 0.47$
...	0.42	0.44	0.50	$q_{0.975,100} = 0.42$	$h_{0.975,100} = 0.46$

More observations in the tail \Rightarrow less noise \Rightarrow more robust estimates

Rockafellar, R.T. and S. Uryasev (2002), "Conditional Value-at-Risk for General Loss Distributions," *Journal of Banking and Finance* 26, 1443-1471.

Problem Formulation

$\ell_i(\mathbf{w}, S^o)$ is the loss of portfolio $\mathbf{w} \in \Omega$ in scenario i of the sample S^o

VaR minimization

$$\min q_{\alpha,N}$$

s.t.

$$\ell_i(\mathbf{w}, S^o) - Mz_i - q_{\alpha,N} \leq 0, \quad i = 1, \dots, N$$

$$\sum_{i=1}^N z_i \leq \lfloor N(1-\alpha) \rfloor$$

$$z_i \in \{0, 1\}, \quad i = 1, \dots, N$$

$$\mathbf{w} \in \Omega .$$

MIP formulation

CVaR minimization

$$\min u + \frac{1}{N(1-\alpha)} \sum_{i=1}^N y_i$$

s.t.

$$\ell_i(\mathbf{w}, S^o) - u - y_i \leq 0, \quad i = 1, \dots, N$$

$$y_i \geq 0, \quad i = 1, \dots, N$$

$$\mathbf{w} \in \Omega .$$

LP formulation

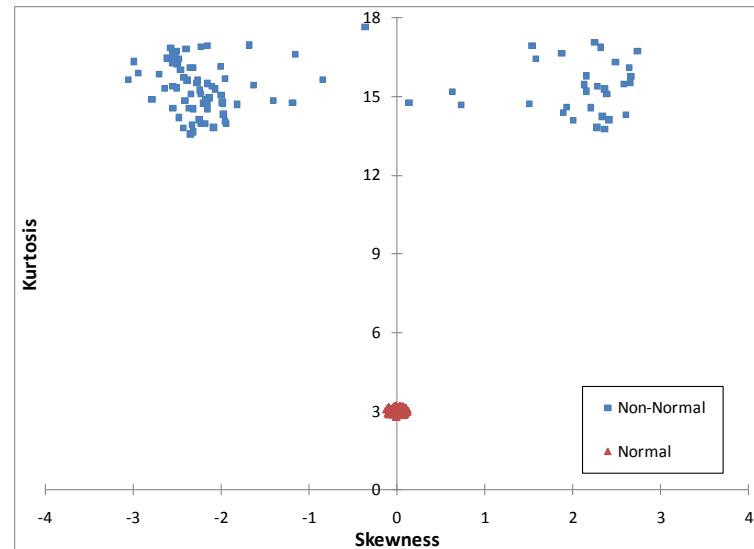
The larger $N(1 - \alpha)$, i.e., # tail observations, the longer the solution time

- Better VaR estimates, in particular, come with high computational cost

Asset Return Scenarios

International stocks selected based on monthly returns from January 2003 to June 2010 (90 months)

- 100 stocks with Normal returns
- 100 stocks with non-Normal returns (skewed, leptokurtic)



Generate return scenarios that match the first four moments and the correlations of the historical stock returns

- One set of 1,000,000 scenarios represents the true return distribution
- 25 sets of 20,000 scenarios for optimization (S^o)

Høyland, K., Kaut, M. and S.W. Wallace (2003), "A Heuristic for Moment-Matching Scenario Generation," *Computational Optimization and Applications* 24, 169-185

Experimental Design

Consider the following parameters

- Multivariate asset return distributions: $\mathcal{F} = \{\text{Normal, Non-Normal}\}$
- Feasible region: $\Omega := \left\{ \mathbf{w} \in \mathbb{R}^J : \sum_{j=1}^J w_j = 1, \quad w_j \geq 0 \text{ for } j = 1, \dots, J \right\}$
- Sample sizes: $\mathcal{N} = \{1000, 5000, 10000, 20000\}$
- Risk measures: $\Pi = \{\sigma^2, \eta_{0.90}, \eta_{0.95}, \eta_{0.99}, \eta_{0.999}\}, P_N = \{s^2_N, h_{0.90,N}, h_{0.95,N}, h_{0.99,N}, h_{0.999,N}\}$

For each $\mathfrak{F} \in \mathcal{F}, N \in \mathcal{N}, b_N \in P_N$

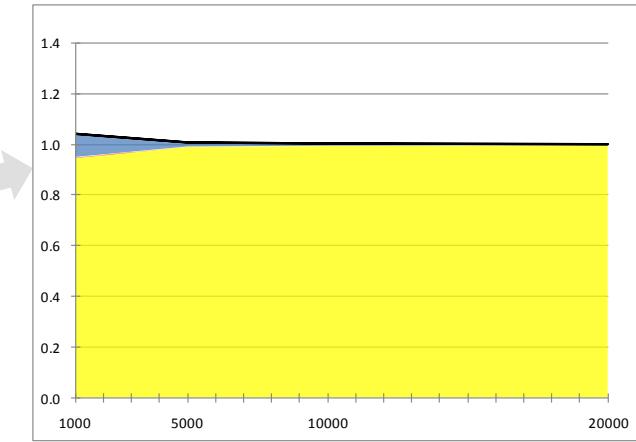
- Find \mathbf{w}^* that minimizes $b_N(\mathbf{w}, S^o)$
- For each $\pi \in \Pi$ and its estimator $p_N \in P_N$
 - Compute optimism = $\pi(\mathbf{w}^*) - p_N(\mathbf{w}^*, S^o)$
 - Compute exploitation = $\mathbb{E}_S[p_N(\mathbf{w}^*, S)] - p_N(\mathbf{w}^*, S^o)$
 - Compute bias = $\mathbb{E}_S[p_N(\mathbf{w}^*, S)] - \pi(\mathbf{w}^*)$

Perform 25 trials with different samples S^o and average the results

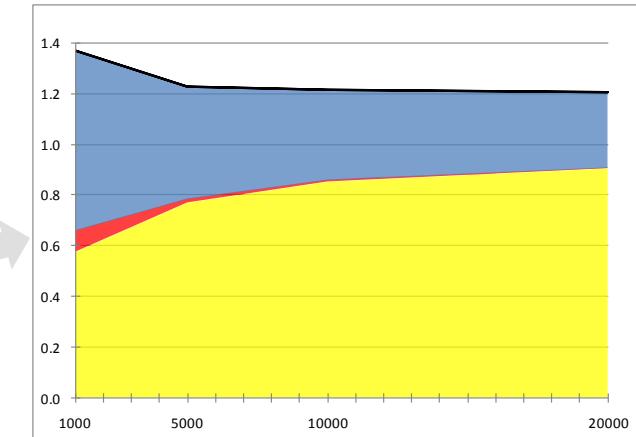
Performance of Natural Estimators

Actual risk (A), exploitation (E) and bias (B) as percentage of true optimum

p_N	N	Normal			Non-Normal		
		A	E	B	A	E	B
s^2_N	1000	102.1	4.7	0.0	104.7	11.0	0.0
	5000	100.4	0.3	0.0	100.9	2.5	0.0
	10000	100.1	-0.1	0.0	100.4	1.4	0.0
	20000	100.0	0.0	0.0	100.0	-0.1	0.0
$h_{0.90,N}$	1000	104.2	9.4	-0.2	106.0	13.3	-0.2
	5000	100.8	1.7	0.0	101.7	4.0	-0.1
	10000	100.4	0.9	0.0	101.2	2.4	0.0
	20000	100.1	0.4	0.0	100.8	1.4	0.0
$h_{0.95,N}$	1000	105.7	13.2	-0.2	108.2	18.4	-0.3
	5000	101.4	2.8	-0.1	102.6	5.9	-0.1
	10000	100.6	1.3	0.0	101.7	3.5	0.0
	20000	100.3	0.6	0.0	101.2	2.2	0.0
$h_{0.99,N}$	1000	111.6	28.2	-0.9	119.2	40.5	-1.2
	5000	103.1	7.1	-0.2	107.0	13.9	-0.2
	10000	101.8	3.7	-0.1	105.6	9.2	-0.1
	20000	101.0	2.3	0.0	104.7	6.9	0.0
$h_{0.999,N}$	1000	114.3	45.2	-4.7	137.2	71.2	-8.6
	5000	110.0	25.3	-1.1	122.9	44.3	-1.6
	10000	107.1	16.8	-0.6	121.6	35.6	-0.8
	20000	104.8	10.5	-0.3	120.7	29.8	-0.3

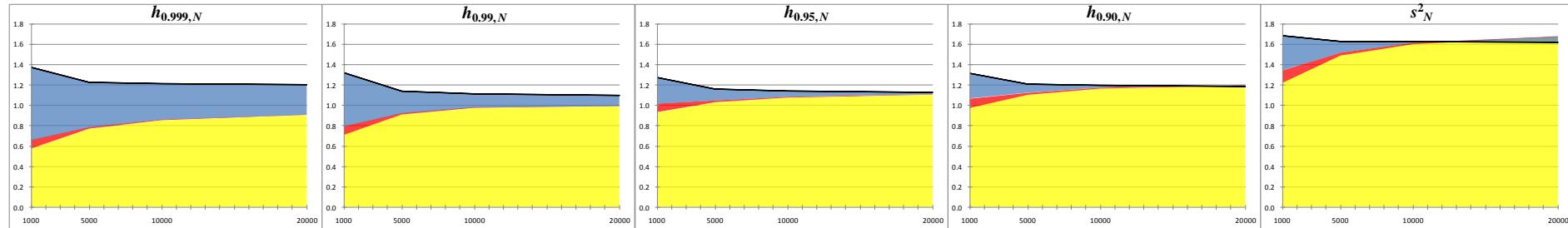


■ Perceived Risk ■ Exploitation ■ Bias (-)

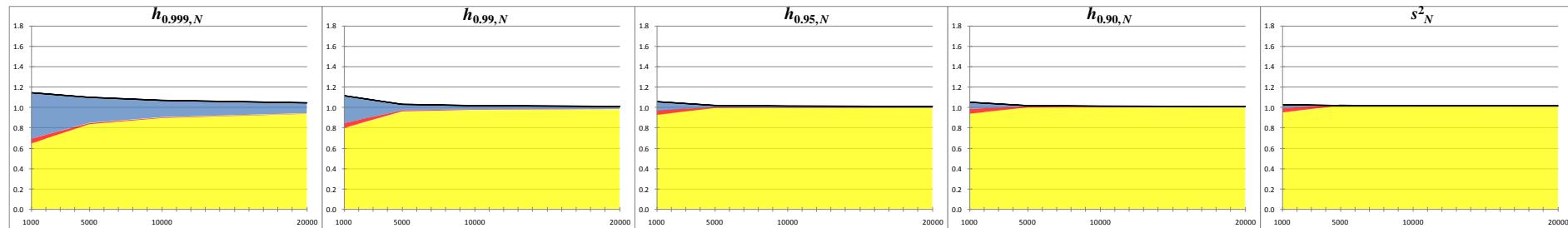


Proxies for CVaR(99.9%)

Non-Normal



Normal



- Why is variance such an effective proxy for CVaR(99.9%) in this case?
 - $\mathcal{L} \sim N(\mu, \sigma^2) \Rightarrow \eta_\alpha = \mu + \hat{\eta}_\alpha \sigma$ where $\hat{\eta}_\alpha$ is CVaR(α) for the $N(0, 1)$ distribution
 - As $\alpha \rightarrow 1$, $\hat{\eta}_\alpha \rightarrow \infty$ and so minimizing σ^2 effectively minimizes η_α

Performance of Proxies

Difference in actual risk (%) relative to natural estimator

N	b_N	π	Normal				Non-Normal					
			σ^2	$\eta_{0.90}$	$\eta_{0.95}$	$\eta_{0.99}$	$\eta_{0.999}$	σ^2	$\eta_{0.90}$	$\eta_{0.95}$	$\eta_{0.99}$	
1000	s^2_{1000}		-	2.8	-0.4	-7.3	-10.2	-	13.8	16.6	19.7	23.0
	$h_{0.90,1000}$		14.8	-	-1.4	-6.3	-8.1	34.5	-	-0.8	-4.2	-4.1
	$h_{0.95,1000}$		15.2	1.7	-	-5.2	-7.2	45.6	1.8	-	-5.2	-6.9
	$h_{0.99,1000}$		24.5	8.4	6.1	-	-2.4	63.0	9.1	6.5	-	-3.5
	$h_{0.999,1000}$		30.7	11.3	8.9	2.6	-	67.5	12.6	9.9	3.4	-
5000	s^2_{5000}		-	5.6	3.2	-0.4	-7.4	-	15.6	19.6	29.2	32.7
	$h_{0.90,5000}$		11.2	-	-0.4	-1.6	-7.2	37.8	-	-0.3	0.2	-1.4
	$h_{0.95,5000}$		9.8	0.6	-	-1.4	-7.2	48.5	1.4	-	-1.5	-5.6
	$h_{0.99,5000}$		10.3	3.0	2.0	-	-6.2	75.2	7.0	4.1	-	-7.3
	$h_{0.999,5000}$		24.4	10.3	9.0	6.8	-	92.7	15.6	12.2	7.8	-
10000	s^2_{10000}		-	5.8	3.8	0.7	-5.1	-	16.3	20.7	31.0	33.9
	$h_{0.90,10000}$		10.5	-	-0.1	-0.7	-5.1	37.2	-	0.1	1.0	-1.4
	$h_{0.95,10000}$		8.8	0.3	-	-0.9	-5.4	48.9	1.1	-	-1.3	-6.2
	$h_{0.99,10000}$		8.3	2.0	1.3	-	-4.9	78.9	6.9	4.1	-	-8.5
	$h_{0.999,10000}$		17.5	8.2	7.1	5.4	-	119.6	19.5	15.8	10.4	-
20000	s^2_{20000}		-	6.1	4.1	1.4	-3.0	-	16.6	21.1	31.8	34.4
	$h_{0.90,20000}$		10.2	-	-0.0	-0.3	-3.2	37.8	-	0.1	1.2	-1.5
	$h_{0.95,20000}$		8.0	0.3	-	-0.5	-3.6	50.2	1.1	-	-1.2	-6.6
	$h_{0.99,20000}$		6.6	1.7	1.0	-	-3.5	83.5	6.9	4.1	-	-9.0
	$h_{0.999,20000}$		12.8	6.2	5.2	3.8	-	139.2	21.6	17.4	11.6	-
∞	σ^2		-	6.2	4.4	2.5	1.6	-	17.5	22.6	37.9	62.1
	$\eta_{0.90}$		10.1	-	0.1	0.7	1.3	35.2	-	0.7	5.4	18.0
	$\eta_{0.95}$		7.4	0.1	-	0.2	0.7	46.7	0.6	-	2.2	11.1
	$\eta_{0.99}$		4.7	0.7	0.2	-	0.2	73.2	4.2	1.7	-	2.6
	$\eta_{0.999}$		3.4	1.7	0.9	0.2	-	108.3	11.2	7.0	2.2	-

Some Observations

Effective proxies are less extreme than the natural estimator

- Use more of the sample data \Rightarrow less noise sensitivity \Rightarrow less exploitation

Benefits of proxies are greater as noise increases

- Smaller sample size, higher quantile level

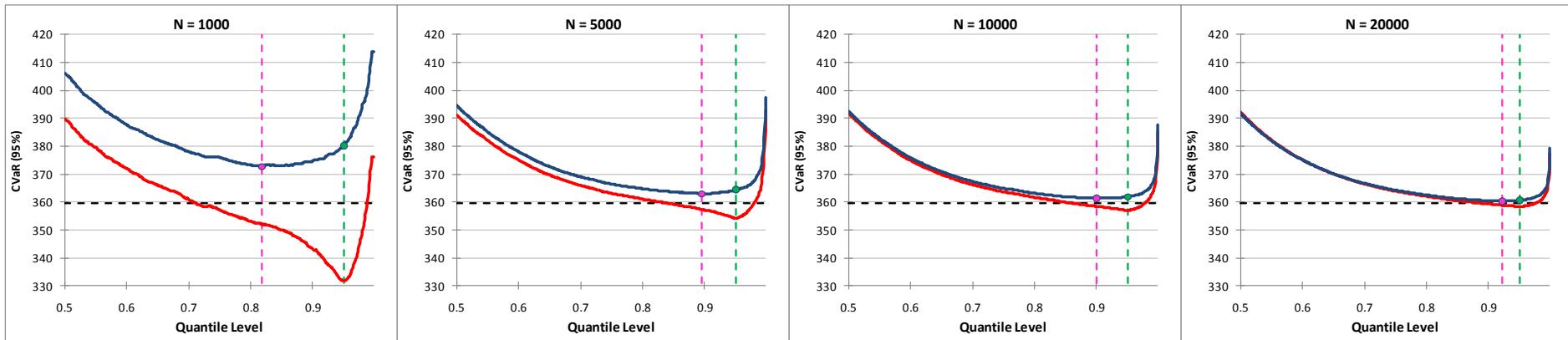
More proxies are effective for Normal returns

- Risk measures are more “substitutable” than for non-Normal returns

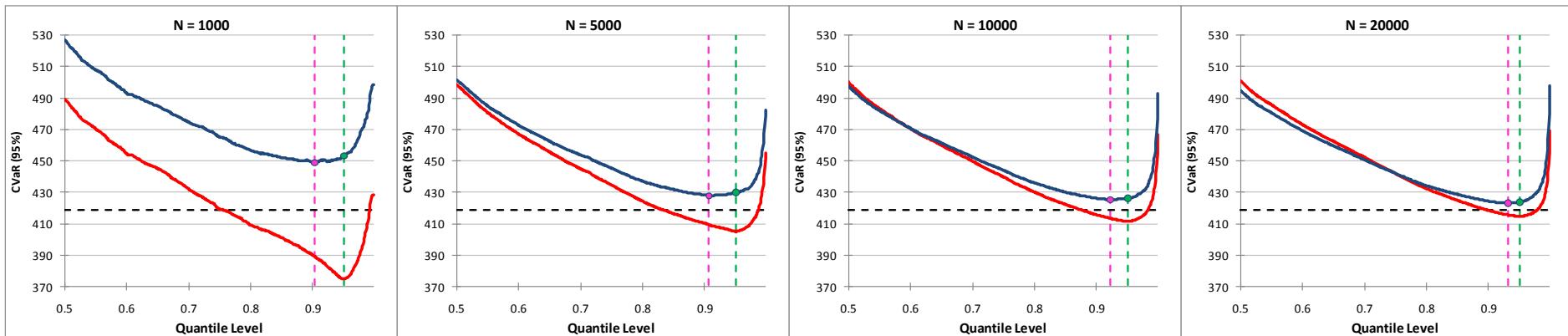
What is the best quantile level for the CVaR proxy?

CVaR Proxy Frontiers for CVaR(95%)

Normal



Non-Normal



Best Quantile Levels for CVaR Proxies

CVaR Level	N	Normal		Non-Normal	
		Best α'	Risk Diff (%)	Best α'	Risk Diff (%)
90%	1000	0.818	-0.822	0.855	-0.172
	5000	0.893	-0.048	0.866	-0.133
	10000	0.874	-0.070	0.870	-0.108
	20000	0.885	-0.010	0.882	-0.038
95%	1000	0.818	-1.961	0.903	-0.941
	5000	0.895	-0.415	0.907	-0.477
	10000	0.900	-0.126	0.922	-0.170
	20000	0.922	-0.070	0.932	-0.172
99%	1000	0.854	-6.569	0.945	-5.299
	5000	0.922	-1.617	0.970	-2.088
	10000	0.940	-0.875	0.977	-1.770
	20000	0.952	-0.525	0.964	-1.750
99.9%	1000	0.867	-8.222	0.963	-7.306
	5000	0.922	-7.311	0.983	-8.202
	10000	0.971	-5.467	0.983	-9.276
	20000	0.973	-3.769	0.982	-9.463

variance is an even better proxy

Trade-off between perceived risk and exploitation is key to proxy performance

- Proxies always have higher perceived risk than the natural estimator
- If the decrease in exploitation offsets the increase in perceived risk then the proxy is effective
- As the sample size increases, exploitation diminishes and perceived risk dominates
 - Most effective proxy approaches the natural estimator

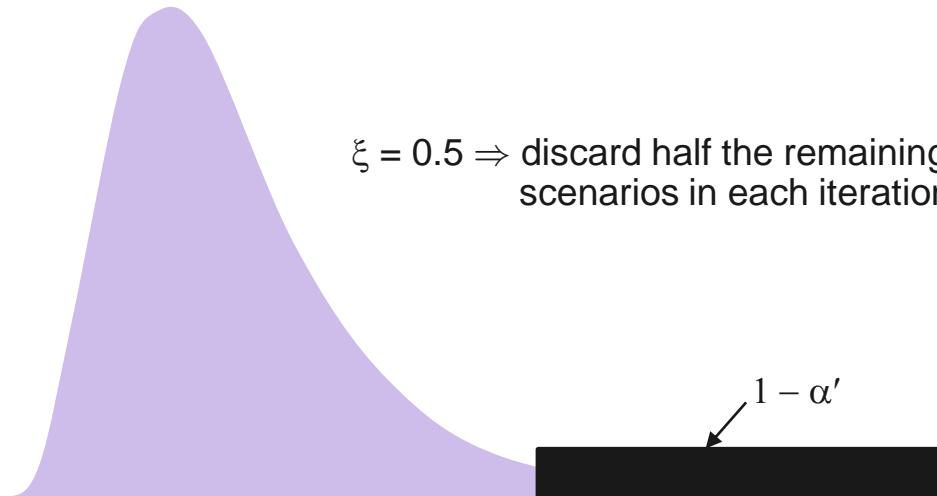


VaR Minimization

Repeat previous experiment for $\text{VaR}(\alpha)$, $\alpha = 90\%, 95\%, 99\%, 99.9\%$

Additional proxies considered for $\text{VaR}(\alpha)$

- $\text{VaR}(\alpha')$ with MIP (30 minute time limit)
- $\text{VaR}(\alpha')$ with heuristic of Larsen et al
 - Minimize $\text{CVaR}(\alpha')$ then iteratively “discard” tail scenarios
 - We report results for $\xi = 1.0$



Larsen, N., Mausser H., and S. Uryasev (2002), “Algorithms for Optimization of Value-at-Risk,” in *Financial Engineering, e-commerce and Supply Chain*, P. Pardalos and V.K. Tsitsirigos (Eds.), Kluwer, Norwell, MA, 129-157.

Performance of Proxies for VaR

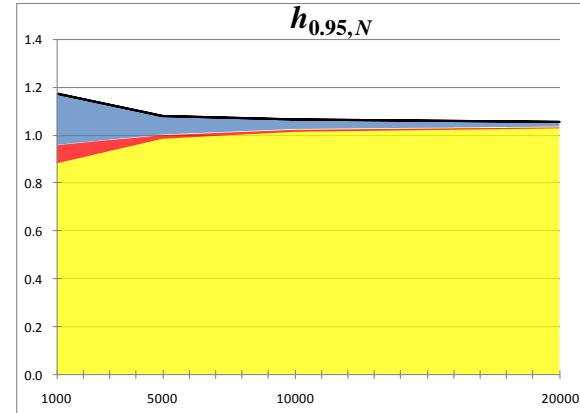
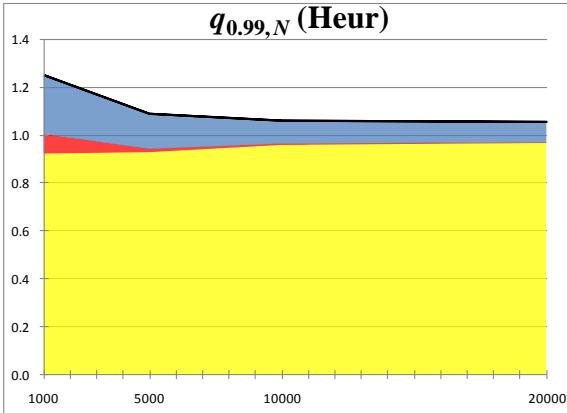
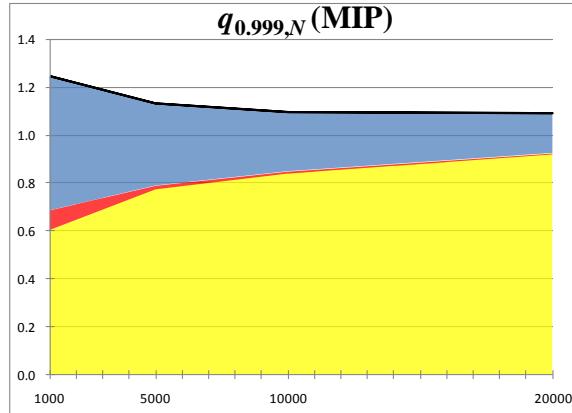
Actual risk as percentage of lowest known risk (best quantile level for proxy)

N	b_N	π	Normal				Non-Normal			
			$\theta_{0.90}$	$\theta_{0.95}$	$\theta_{0.99}$	$\theta_{0.999}$	$\theta_{0.90}$	$\theta_{0.95}$	$\theta_{0.99}$	$\theta_{0.999}$
1000	$q_{\alpha',1000}$ (MIP)	110.8 (95)	109.9 (95)	109.9 (95)	110.4 (99)	114.1 (95)	112.8 (99)	114.4 (99)	124.7 (99.9)	
	$q_{\alpha',1000}$ (Heur)	106.9 (90)	105.7 (90)	105.4 (90)	105.8 (90)	111.4 (90)	107.8 (90)	111.1 (90)	121.9 (95)	
	s^2_{1000}	113.6	108.2	104.3	102.7	114.6	116.2	130.1	155.0	
	$h_{\alpha',1000}$	105.4 (90)	104.4 (90)	104.2 (90)	104.6 (90)	109.8 (90)	105.8 (90)	108.1 (90)	117.2 (95)	
5000	$q_{\alpha',5000}$ (MIP)	108.2 (99)	106.2 (99)	105.2 (99)	105.2 (99)	110.7 (95)	108.1 (99)	106.6 (99)	113.2 (99)	
	$q_{\alpha',5000}$ (Heur)	102.0 (90)	101.3 (90)	101.4 (90)	101.8 (90)	107.3 (90)	102.6 (90)	102.9 (95)	108.8 (95)	
	s^2_{5000}	113.0	107.5	103.5	101.9	112.3	113.6	126.5	149.8	
	$h_{\alpha',5000}$	101.6 (90)	100.9 (90)	101.1 (90)	101.5 (90)	107.0 (90)	102.2 (90)	102.4 (95)	107.9 (99)	
10000	$q_{\alpha',10000}$ (MIP)	108.9 (99)	106.4 (99.9)	104.9 (99.9)	104.7 (99.9)	118.1 (99)	110.5 (99)	107.9 (99.9)	109.8 (99.9)	
	$q_{\alpha',10000}$ (Heur)	101.4 (90)	100.7 (90)	100.8 (90)	101.0 (95)	106.4 (90)	101.9 (90)	101.8 (95)	105.9 (99)	
	s^2_{10000}	112.7	107.3	103.4	101.7	112.2	113.6	126.6	149.7	
	$h_{\alpha',10000}$	101.2 (90)	100.5 (90)	100.6 (95)	100.8 (95)	106.3 (90)	101.7 (90)	101.5 (95)	106.2 (99)	
20000	$q_{\alpha',20000}$ (MIP)	108.1 (99.9)	105.5 (99.9)	104.0 (99.9)	103.6 (99.9)	127.1 (99.9)	115.6 (99.9)	108.5 (99.9)	109.3 (99.9)	
	$q_{\alpha',20000}$ (Heur)	101.0 (90)	100.3 (90)	100.4 (95)	100.5 (95)	106.0 (90)	101.5 (90)	101.0 (95)	105.3 (99)	
	s^2_{20000}	112.6	107.3	103.3	101.6	112.1	113.5	126.3	149.3	
	$h_{\alpha',20000}$	100.9 (90)	100.2 (90)	100.3 (95)	100.4 (95)	105.9 (90)	101.4 (90)	100.9 (95)	105.0 (99)	

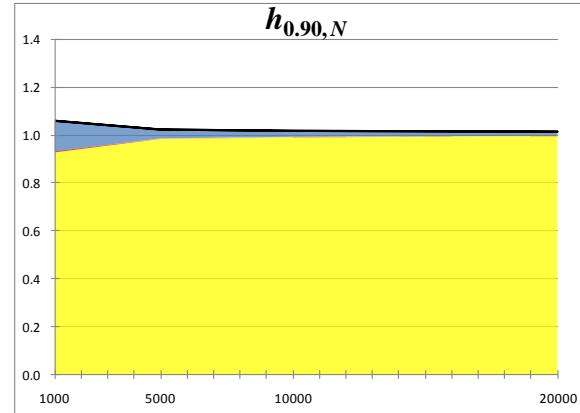
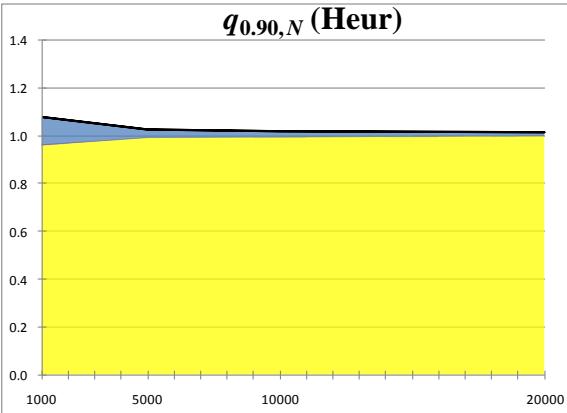
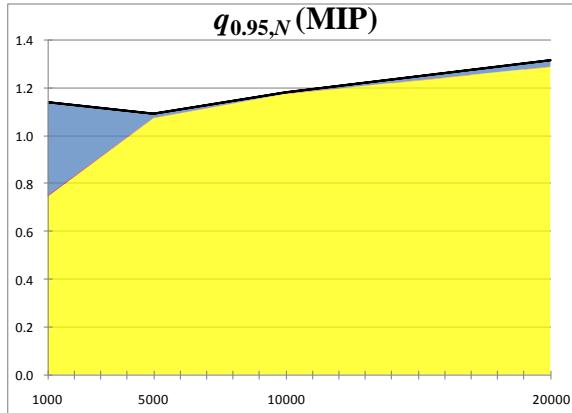
- MIP proxies usually do better when $\alpha' \geq \alpha$ (smaller $N(1 - \alpha) \Rightarrow$ easier to solve)

Proxies for VaR

VaR(99.9%), Non-Normal



VaR(95%), Non-Normal



Best Quantile Levels for CVaR Proxies

VaR Level	N	Normal	Non-Normal
90%	1000	0.726	0.691
	5000	0.766	0.685
	10000	0.736	0.681
	20000	0.734	0.682
95%	1000	0.810	0.815
	5000	0.855	0.808
	10000	0.855	0.811
	20000	0.863	0.814
99%	1000	0.854	0.924
	5000	0.910	0.940
	10000	0.935	0.955
	20000	0.946	0.953
99.9%	1000	0.867	0.958
	5000	0.922	0.975
	10000	0.972	0.980
	20000	0.969	0.979

For Normal distributions:

$\text{CVaR}(\alpha') = \text{VaR}(\alpha)$ if α' satisfies

$$\frac{\phi(Z_\alpha)}{1-\alpha'} = Z_\alpha$$

$\text{VaR}(90\%) = \text{CVaR}(75.44\%)$

$\text{VaR}(95\%) = \text{CVaR}(87.45\%)$

$\text{VaR}(99\%) = \text{CVaR}(97.42\%)$

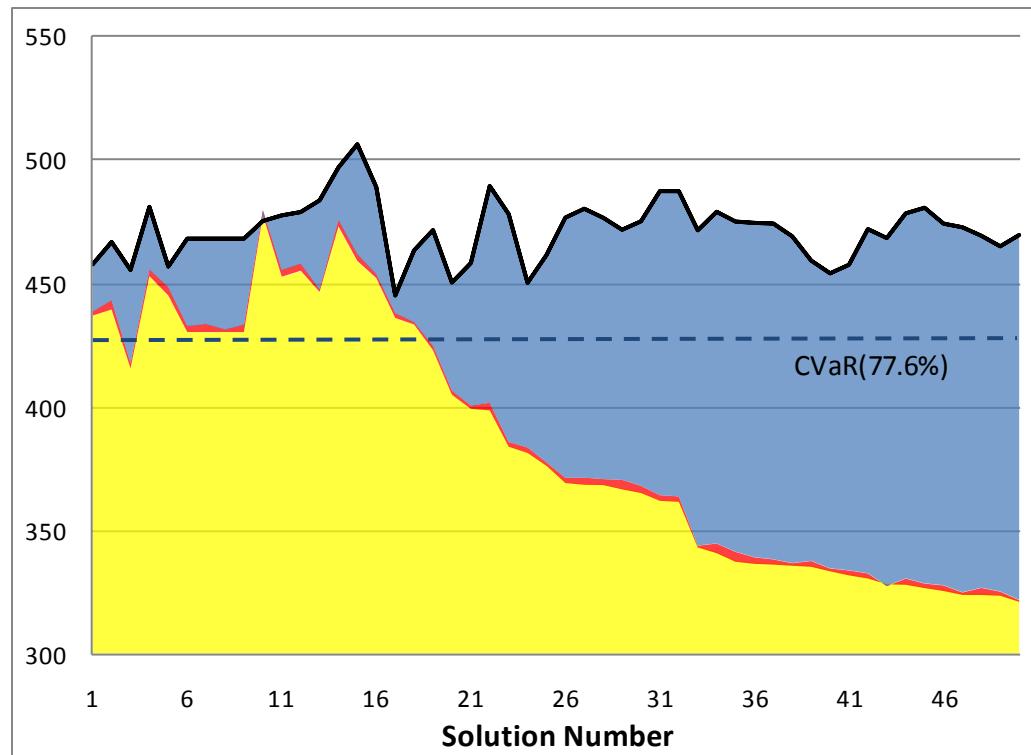
$\text{VaR}(99.9\%) = \text{CVaR}(99.74\%)$

variance is an even better proxy

MIP Performance for VaR(95%)

Non-Normal, $N = 1000$ (one trial)

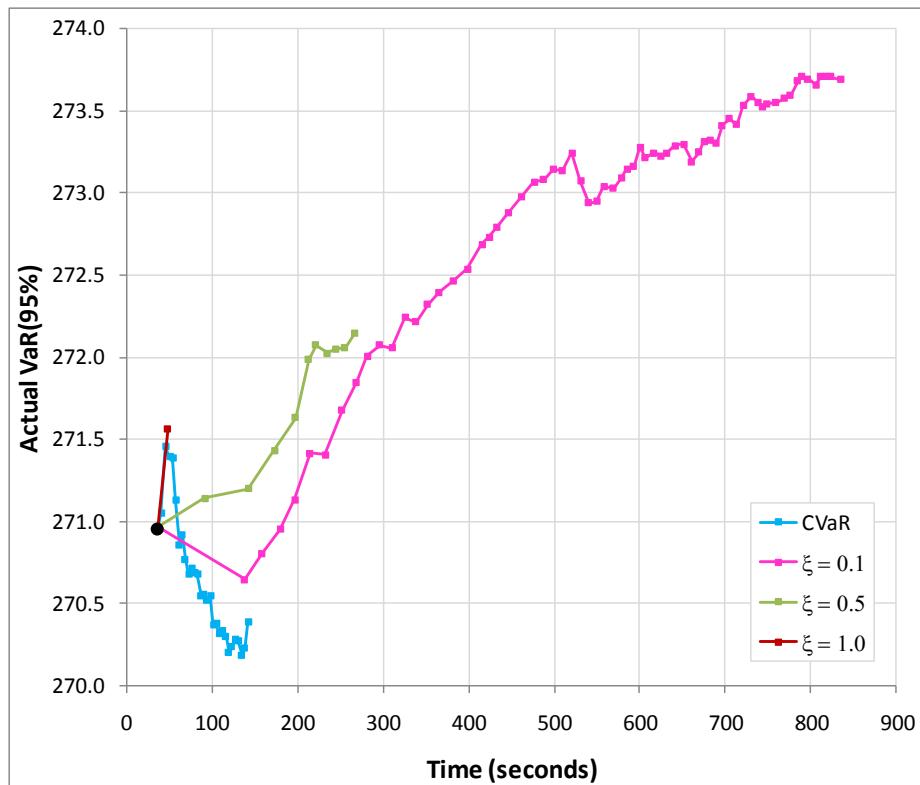
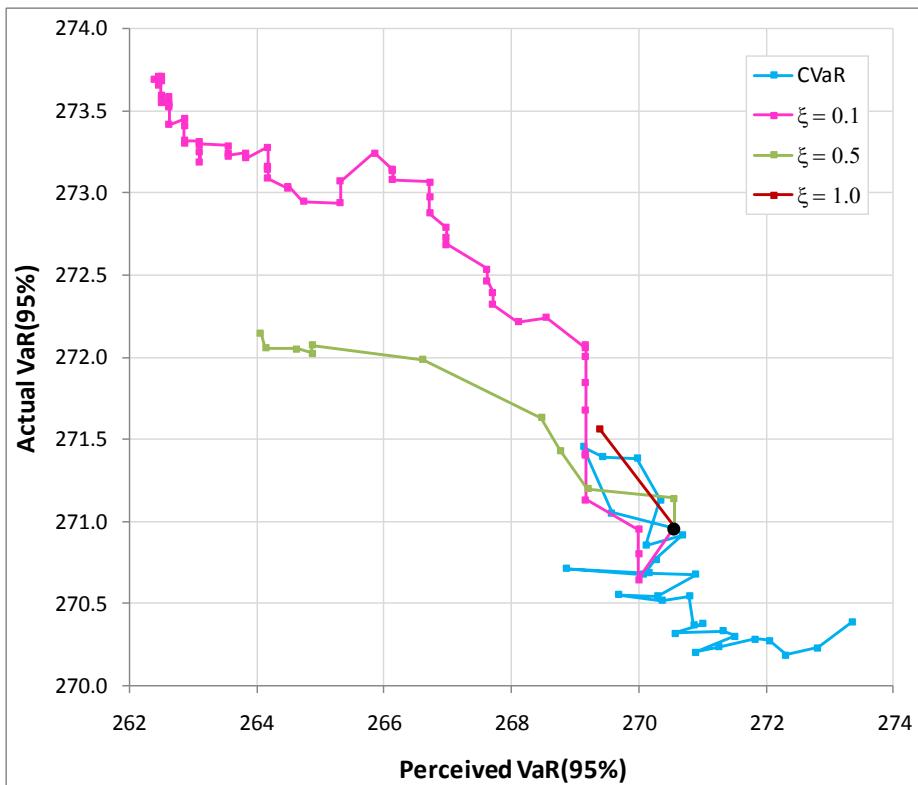
- Perceived risk progressively decreases but actual risk does not
- After 30 minutes, best MIP solution does not match the CVaR proxy (10 seconds)



Heuristic Performance for VaR(95%)

Normal, $N = 20000$ (one trial)

- Perceived risk progressively decreases but actual risk does not
- CVaR proxy with quantile level stepsize of 0.5% performs better



Practical Issues

CVaR is an effective proxy but need to find the appropriate quantile level

- Solve a series of CVaR problems and evaluate solutions out-of-sample

Fortunately, the CVaR problems can be solved quickly

- Changing the quantile level only affects the objective function \Rightarrow solution stays primal feasible \Rightarrow warm start is effective
- Our results (CPLEX 12.2) suggest that, on average
 - subsequent CVaR problems take 10% as long as the initial problem
 - 5 CVaR problems can be solved per iteration of the VaR heuristic

Integrating out-of-sample evaluation into the optimization procedure can speed up the “brute force” approach

- Bisection, stepsize refinement, etc.

Questions?

