Optimal Portfolio Construction with Estimation Error, Non-normal Returns, and Large Numbers of Assets

John Birge

University of Chicago

Booth School of Business

General Theme

- Portfolio optimization is difficult due to:
  - Consistency
  - Many parameters to estimate
  - Non-stationarity
  - Non-normality
  - …
- Optimization models are driven to extremes and naturally focus on "rare events" that can create problems
- Most convergence results rely on asymptotic results and constants that are difficult to estimate

CHICAGO BOOTH

Example: Financial Portfolio Optimization

Quadratic program (Markowitz Portfolio):

find investments x=(x(1),…,x(n)) to

$$\min x^T Q x$$

$$\text{s.t. } r^T x = \text{target},\ e^T x=1$$

where Q and r are typically estimated from historical data.

Correlations from University of Michigan CIO:

	DomCommor	SmallCap	InteCommon	EmerMarkets	AbsoluteRetu	VentCap	RealEst	Oil and Gas	Commodities	FixedIncome	IntFixedInc	Cash
DomCommor	1	0.79	0.58	0.56	0.6	0.44	0.25	0.01	-0.3	0.43	0.2	0.27
SmallCap	0.79	1	0.48	0.61	0.65	0.56	0.24	0.01	-0.05	0.31	0.1	0.08
InteCommon	0.58	0.48	1	0.37	0.45	0.25	0.38	-0.04	-0.17	0.35	0.55	0.23
EmerMarkets	0.56	0.61	0.37	1	0.3	0.3	0.07	-0.19	-0.07	-0.07	0.1	0.04
AbsoluteRetu	0.6	0.65	0.45	0.3	1	0.35	0.2	-0.2	0.11	0.35	0.25	0.45
VentCap	0.44	0.56	0.25	0.3	0.35	1	0.21	-0.02	-0.18	0.19	0.15	0.14
RealEst	0.25	0.24	0.38	0.07	0.2	0.21	1	0.08	-0.53	0.15	0.2	0.37
Oil and Gas	0.01	0.01	-0.04	-0.19	-0.2	-0.02	0.08	1	0.54	-0.18	-0.3	-0.07
Commodities	-0.3	-0.05	-0.17	-0.07	0.11	-0.18	-0.53	0.54	1	-0.3	-0.08	-0.13
FixedIncome	0.43	0.31	0.35	-0.07	0.35	0.19	0.15	-0.18	-0.3	1	0.55	0.67
IntFixedInc	0.2	0.1	0.55	0.1	0.25	0.15	0.2	-0.3	-0.08	0.55	1	0.1
Cash	0.27	0.08	0.23	0.04	0.45	0.14	0.37	-0.07	-0.13	0.67	0.1	1

Results from Optimization

	Amt. to invest
DomCommon	-54079107483
SmallCap	-17314640180
InteCommon	-7098209713
EmerMarkets	21285151081
AbsoluteReturn	65911278496
VentCap	3346118938
RealEst	-68300117028
Oil and Gas	66227880617
Commodities	-1.04264E+11
FixedIncome	-72656761796
IntFixedInc	1.17885E+11
Cash	49057530702
Return	0.099999487
Variance	-1.64591E+19

What happened here?

Problems in Markowitz Model

- Consistent time series
  - Correlations from different time series may not yield PD covariance matrices
  - Caution for general parameter estimates
- Number of Correlation Parameters
  - For *n* assets, *n(n-1)/2* correlations to estimate
  - Chances of estimation error increase rapidly in *nru*

Problem Statement

- Large problems with *n* variables and *m* constraints/objective coefficients lead to (at least) *mn* estimates
- Probability of significant deviation from mean values increases rapidly in *mn*
- Deviant estimates drive optimal solutions
- Non-normal returns further exacerbate issues
- ***How can we construct large models that yield consistent results with high probability?***

The General Questions

- Consider the basic problem (stochastic program):

$$\text{Min}_{x \varepsilon X} \; E_{\xi}[f(x,\xi)] \qquad\qquad (P)$$

- Suppose the only information for ξ is through observations: $\xi^1,\ldots,\xi^\nu$
- Typical empirical case:

$$Min_{x\varepsilon\, X}\,(1/\,\nu)\sum_{i=1}^{\nu}\; f\,(x,\xi^i)$$

What is this in relation to solution x^* to (P)?

- What are the best ways to use those observations?

Observations: The Good News

- Asymptotic distribution of optimal solution of sampled problem is:
  - Sometimes multivariate normal
  - Sometimes projection of multivariate normal onto constraints
  - Sometimes an atom at a single point
- Questions for large data sets:
  - When do we start to observe the asymptotic behavior?
  - How big must ν (no. of samples) be?

More Good News

Goal: *Universal Confidence Sets* (e.g., Pflug (2003), Vogel (2008))

$$P\{|E_\xi[f(x^\nu,\xi) - f(x^*,\xi)]| \geq \epsilon\} \leq \alpha_1 e^{-\beta_1 \nu}.$$

and, if x^* is unique,

$$P\{\|x^\nu - x^*\| \geq \epsilon\} \leq \alpha_0 e^{-\beta_0 \nu}.$$

- Possible (sometimes explicit), e.g., Dai, Chen, JRB (2000)
- Can this be used with only empirical observations?

Summary and Questions

- Asymptotic imply that confidence intervals are possible
- Universal bounds indicate exponential convergence to an optimal solution

Questions: 1. When do asymptotic properties appear? (Size of the constants?)

2. What are the effects of dimension? of multiple uncertainties? of constraints?

3. Are there better ways to use samples and, if so, when?

Form of Examples: Mean-Risk

Objective is composed of risk and return:

E[f(x,w)] = - exp.return(x) + risk(x)

For portfolios: -mean + risk-aversion_constant*variance

For uncertainty, sometimes only in the return, sometimes only in risk and sometimes in both – (this can effect convergence)

Example Problem

- Consider the following problem:

$min_x \; E_\xi[\, -\xi^T x + \varepsilon \,||\, x \,||_1 \,]$

$s.\ t.\ -1 <= x <= 1$

where $||\, .||_1$ is the 1-norm (so equivalent to a linear program) and $E[\xi]=0$.

The optimal solution should be $x^*=0$.

How long to achieve limiting distribution?

How long will it take a sample solution to approach x^* exponentially? i.e., when does Log $(P\{||x^\nu - x^*|| \geq \varepsilon\}$ decrease linearly?

Sample Problem

- Assume that $\xi_j \sim N(0,1)$ for all j,

the solution is $x^\nu{}_j = 0$ if $|\xi_j| \leq \varepsilon$, and $\pm$ 1 o.w.

So, $P\{||x^\nu - x^*|| \geq 1\} = P\{|\ x^\nu{}_j\ | \geq 1$, some j$\}$

$= P\{|\xi_j| \geq \varepsilon$, some j$\} = 1-(1-2\Phi(-\ \varepsilon\ \nu^{0.5}))^n$

where Φ is the standard normal c.d.f.

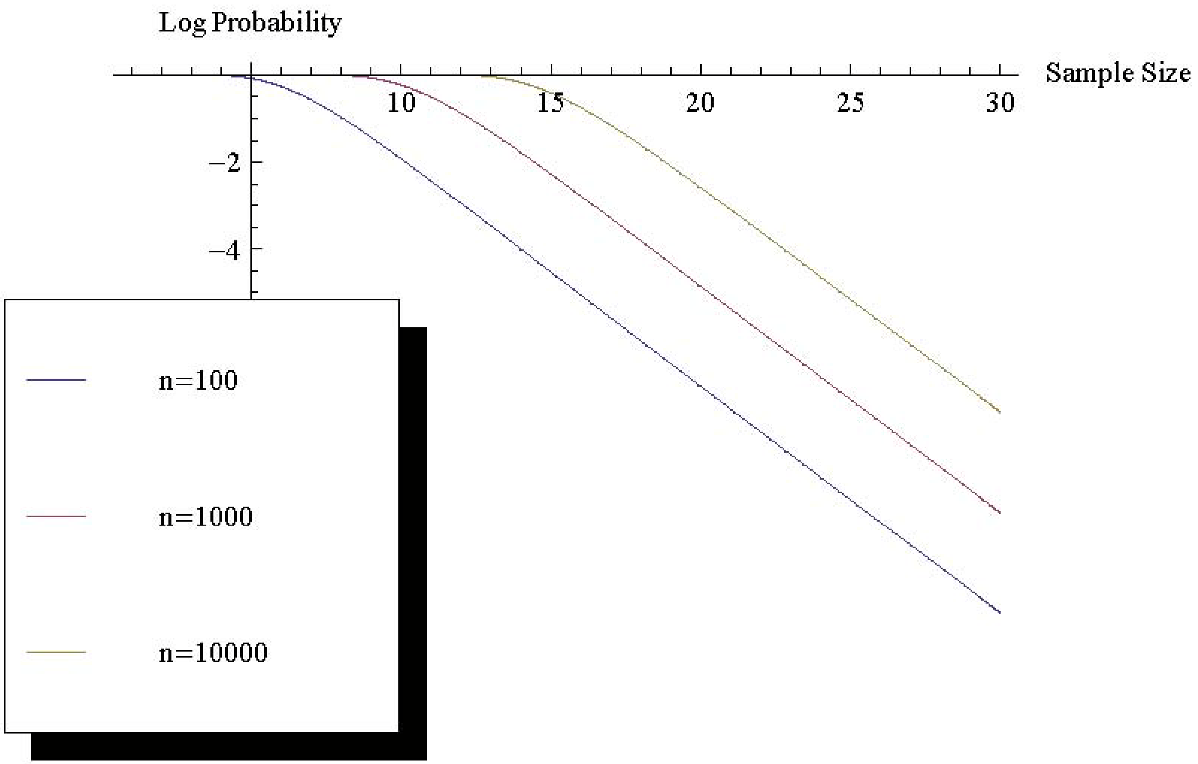
Note: already normal

When is Log(P(error≥1)) linear in ν?

What is the effect of dimension? (Note n)

Results

Log (P(error ≥1)) v. sample size (v)



Observations

- Some delay in approach to exponential error decrease with dimension
- Increase in the delay (size of the constants in the universal bound) is less than linear in dimension (in fact, less than linear in Log of dimension)
- Same kinds of effects for objective
- Good results but could they be even better? Can we reduce the effect of the dimension?

How Can We Reduce the Required Number of Samples?

- Use of sub-samples or batch mean (e.g., Mak, Morton, Wood (99))
- Suppose that we divide the ν samples into k batches of ν/k each, let ξ^{ν}_{i} be the mean of batch $i=1,\ldots,k$, then solve with ξ^{ν}_{i} to obtain x^{ν}_{i}
- Let $x^{\nu,k}=(1/k)\sum_{i=1}^{k} x^{\nu}_{i}$
- When does this perform better than a single sample?
- In particular, how much better in the worst case?
- How does this relate to known portfolio results?

Error Estimates for Portfolios

For sample mean μ^{ν} and sample variance Σ^{ν} with n samples

$x^{\nu}=(\Sigma^{\nu})^{-1}(\mu^{\nu})(\nu\text{-n-2})/\nu$ is an unbiased estimator of x^{*} (for unconstrained case with risk-free asset)

Objective estimate squared is $\chi^{2}{}_{n}(\nu(\mu\Sigma^{-1}\mu))\ )(\nu\text{-n-2})^{2}/\nu^{3}$ with mean: $(n)(\nu\text{-n-2})/\nu^{2} + (\mu\Sigma^{-1}\mu)\ )(\nu\text{-n-2})/\nu$

Note: dependence on n;

With batches:

Variance of $x^{\nu,k}$ is $(n)(\nu\text{-n-2})/\nu^{2} + (\mu\Sigma^{-1}\mu)\ (\nu\text{-n-2-k}\nu)/(\text{k}\nu)$

(assuming independence)

But, sufficient batching can reduce the variance in the estimate of $x^{\nu,k}$ without increasing the number of samples

Result for Sub-sample Batch Optimization – Just Mean Estimate

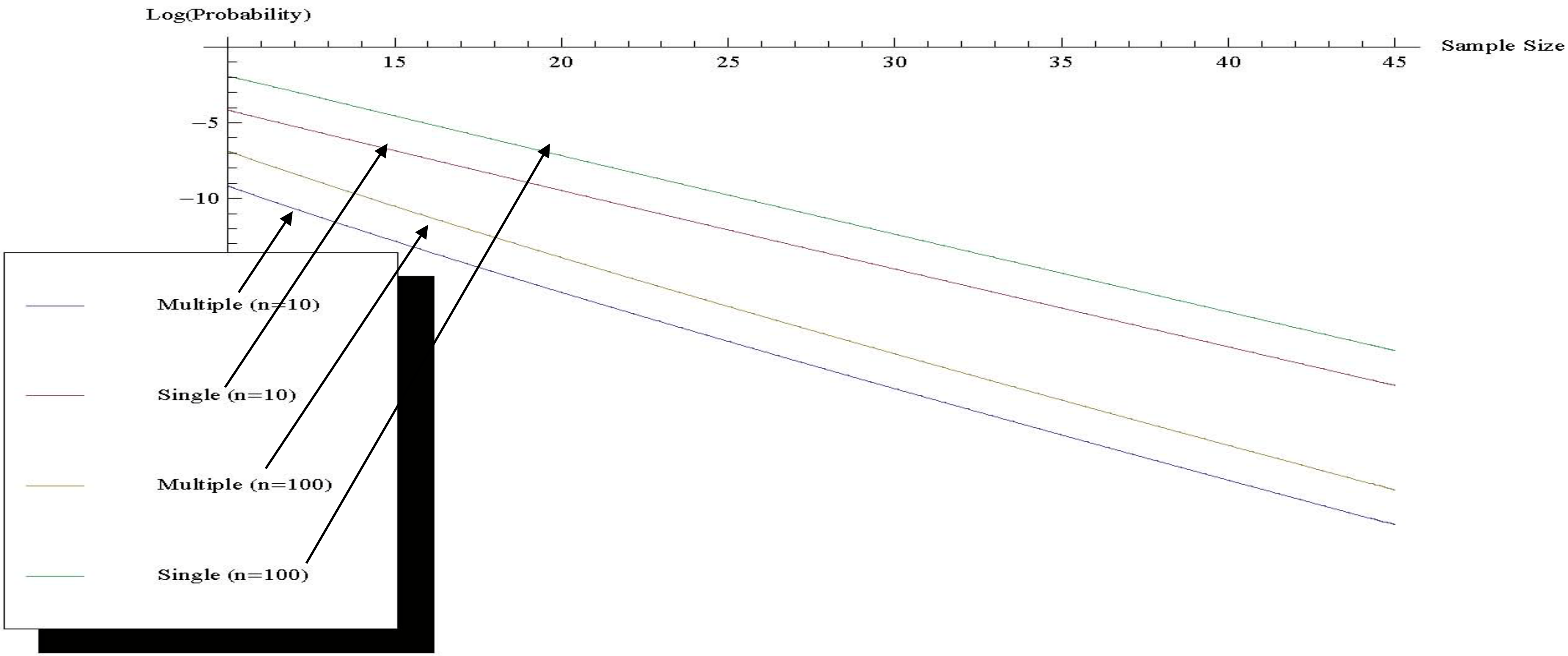
- What is the chance that one component in the decision variable is far off?

$$P\{\|x^{\nu/K,K} - x^*\|_\infty \geq 1\} \;\leq\; P\{|x_j^{\nu,i}| \geq 1, \forall i = 1,\ldots,K; \text{for some } j \in \{1,\ldots,n\},\}$$
$$= \; 1 - (1 - (2\Phi(-\gamma(\nu/K)^{0.5}))^K)^n,$$

- Now, decreased dependence on *n*

CHICAGO BOOTH

Results for Batch/Single Samples



Observe: more improvement as ν ↑ (from 4 to 9 orders of magnitude)

What about Effects of Uncertainty in Risk?

- Example:

$$\min_{\|x\|_2 \leq 1} E[-\xi^T x + \frac{\gamma}{2}\|x\|_2^2],$$

- Now, ξ and γ are random

Suppose ξ_j ~N(0,1); γ~ N(1,1)

- Unconstrained solution:

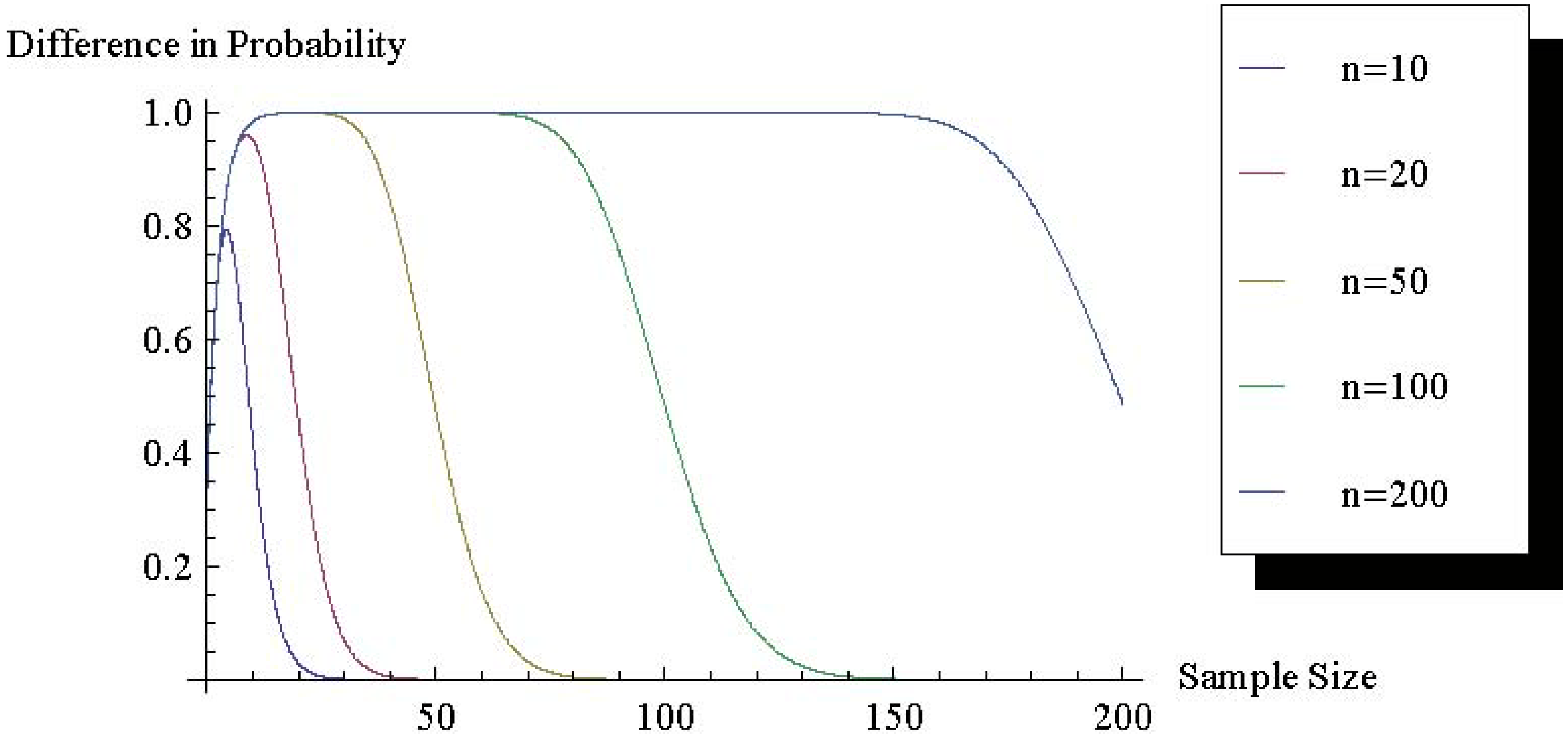
Error in solution in 2-norm is χ^2 under asymptotic distribution

True error in solution is given by:

$$\frac{1}{\|x^{\nu,u} - x^*\|_2^2} \sim F(1, n, \nu),$$

where F is the non-central F-ratio distribution

How Many Samples before the Error Approaches Asymptotic Distribution?



Observations

- Convergence now is much slower than in the case with just stochastic returns
- Convergence delay to the asymptotic distribution is almost linear in dimension
- Asymptotic distribution for the objective is again similar
- Asymptotic distribution for the general portfolio problem with multiple variance estimates (and inverse Wishart distribution) is even worse

Full Portfolio Examples

- General form:

$$\min_{x \in X} -\bar{r}^T x + \frac{\gamma}{2} x^T \Sigma x.$$

requires estimation: e.g., using sample estimates as:

$$\min_{x \in X} -\hat{r}^T x + \frac{\gamma(\nu - n - 2)}{2\nu} x^T \hat{\Sigma} x.$$

and (ν-n-2)/ν term makes solution un-biased with no constraints (e.g., Kan and Zhou (2007))

Questions to Consider

- Does the use of sub-sample/batch optimal solutions improve convergence?
- How do the constraints affect the performance of the batch solution approximations?
- What is the effect of dimension in these problems?

Simulation Setup

For these results, we suppose $n = 10$, $\nu = 500$, and $K = 10$ and let $\gamma = 1$, $\mu = 0.2e$, where $e = (1, \ldots, 1)^T$, and $\Sigma = 0.05 * I$, where I is an identity matrix. We present the results from 1000 simulation runs for three different sets, X, corresponding to increasing ranges on x: $[0,1]^{10}$, $[-1,2]^{10}$, and $[-5,10]^{10}$. The results are compared relative to the optimal solution $x^* = 0.4e$ in terms of $\|x^\nu - x^*\|/\|x^*\|$ and optimal objective value $z^* = -\bar{r}^T x^* + \frac{1}{2} x^{*T} \Sigma x^* = -0.04$ in terms of $(-\bar{r}^T x^\nu + \frac{1}{2}(x^\nu)^T \Sigma x^\nu - z^*)/(-z^*)$.

Observe: histograms of relative errors in solutions and losses in objective

$X=[0,1]^{10}$

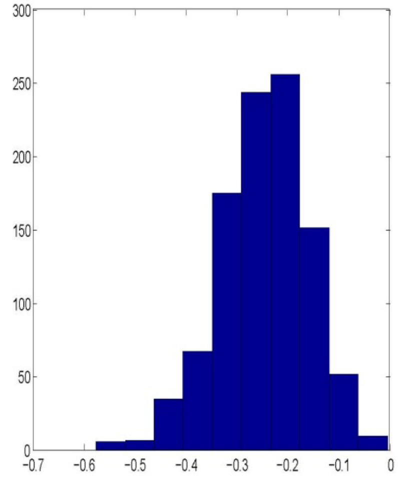
Relatives differences:

Batch better: 1000/1000

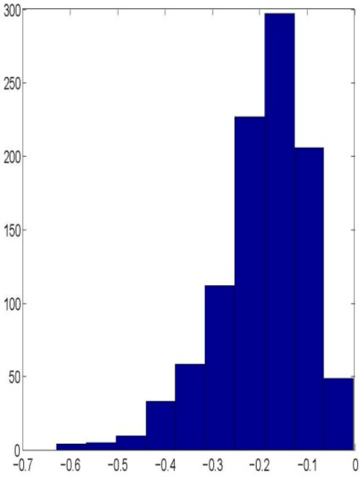
Avg. Sol. Dist. Diff. : -25%

Avg. Obj. Diff.: -19%

Solution



Objective



$X=[-1,2]^{10}$

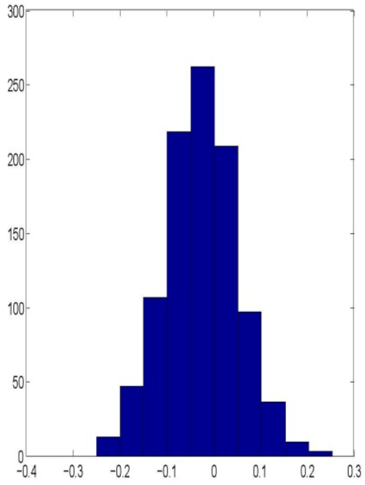
Relatives differences:

Batch better: 638/1000

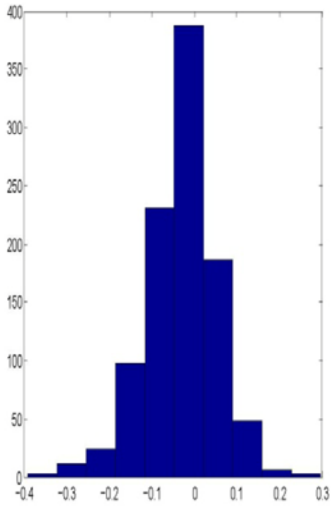
Avg. Sol. Dist. Diff. : -3%

Avg. Obj. Diff.: -3%

Solution



Objective



CHICAGO BOOTH

$X=[-5,10]^{10}$

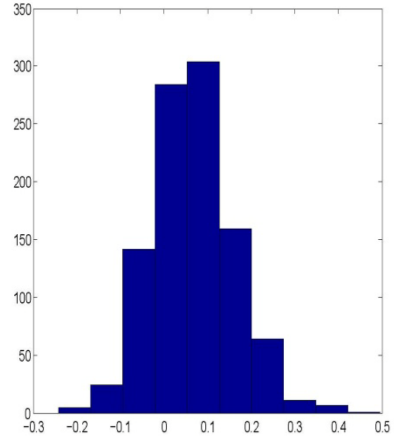
Relatives differences:

Batch better: 231/1000

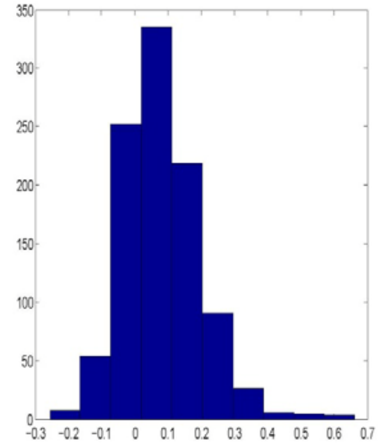
Avg. Sol. Dist. Diff. : +7%

Avg. Obj. Diff.: +8%

Solution

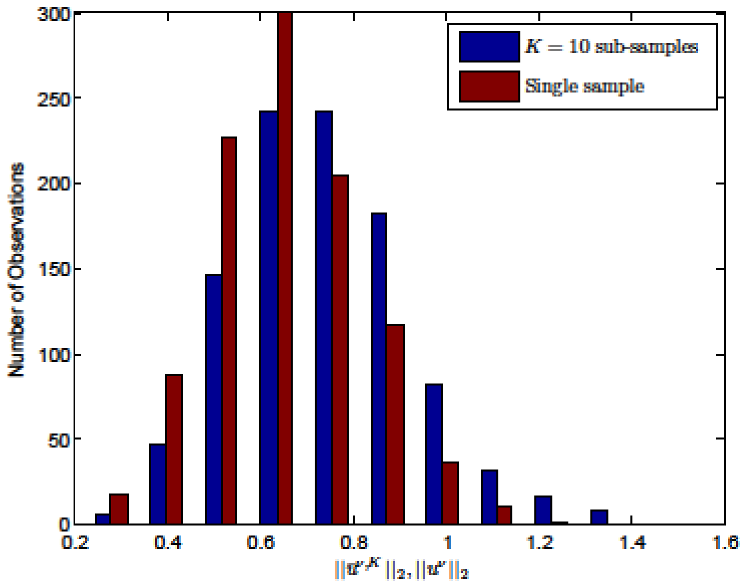


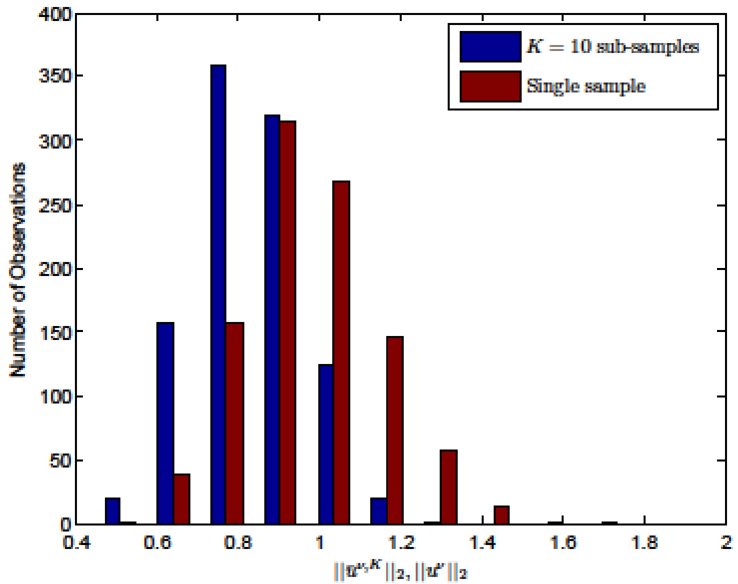
Objective



F

Dimension Effect: X=[-1,2] Relative Distance from Optimum

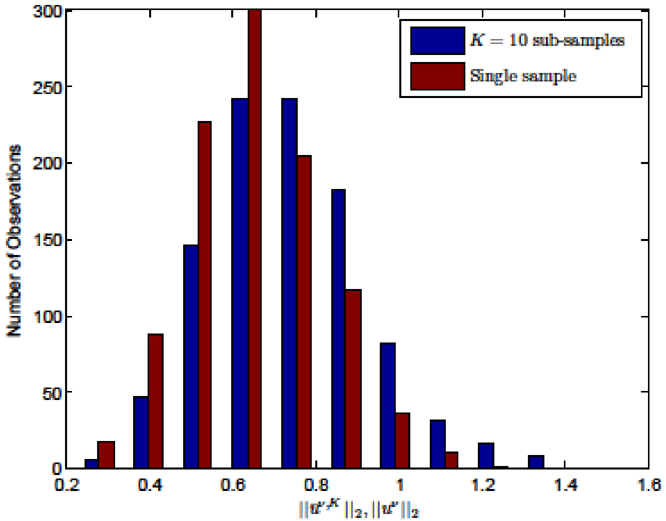




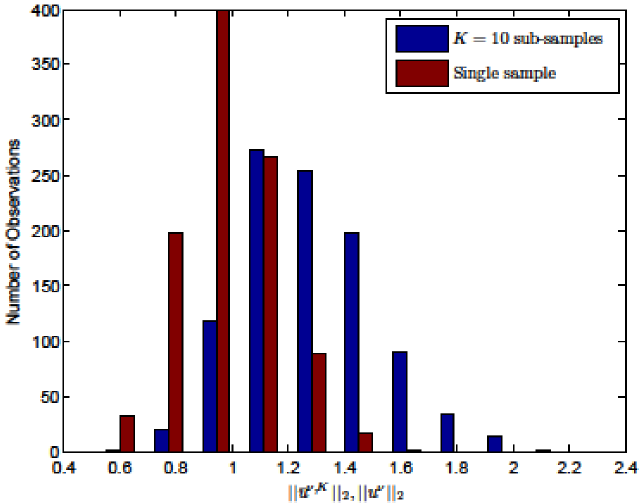
n=10

n=20

Distance Effect: X=[-5,10]
Relative Distance from Optimum



n=10



n=20

Observations on Portfolios

- Batch approach improves when constraints can bind the sample solutions
- The batch improvement is significant when constraints are relatively tight (but still more than 3 standard deviations from optimum)
- Batch can improve without constraints (but not so much in low dimensions ~10)

Effects of Dimension

- Increase dimension from 10 to 20:

Range: $X=[0,1]^{20}$

- 982/1000 better with batching (cf. 1000/1000)
- Average improvement: 15.4% in objective (cf. 19.3%)

$X=[-1,2]^{20}$

- 840/1000 better with batching (cf. 638/1000)
- Average improvement: 8.5% in objective (cf. 3.0%)

$X=[-5,10]^{20}$

- 39/1000 better with batching (cf. 231/1000)
- Average improvement: -22% in objective (cf. -8%)

Observation: there is a “sweet spot” in constraints where batching is especially effective

Observations on Portfolios

- Batch approach improves when constraints can bind the sample solutions
- The batch improvement is significant when constraints are relatively tight (but still more than 3 standard deviations from optimum)
- Batch can improve without constraints (but not so much here)
- Batching favored by medium constraint level

Alternative Responses

- Factor methods (but still need estimation for factors and coefficients)
- Robust optimization
- Robust estimation
- Bayesian updating/Estimation with non-data information
- Simple rules
- Non-normal distribution assumptions

Robust Optimization

Idea: Suppose an uncertainty set around the estimated data

Optimize over the worst case in the uncertainty set

Example: $(r, V) \in \mathcal{R} \times \mathcal{W}$

Min (Max$_{(r,V)\in\mathcal{R}\times\ \mathcal{W}}\, x^T\, V\, x$

s.t. $r^T\, x \geq r^*$, $e^T\, x=1$ $(x\geq 0)$)

Challenges in Robust Optimization

- Choice of uncertainty set
  - Usually set outside of model (ad hoc)
  - If defined as confidence interval based on observations, must grow larger with problem size to avoid aberrant solutions
- Solution structure
  - Solution avoids assets with large uncertainty sets (i.e., sets to 0)
  - May yield lack of diversification

Bayesian and Non-data Procedures

Assume some prior on structure of returns and co-variances (e.g., Black-Litterman)

- Use CAPM equilibrium
- All prices are consistent
- Weights on all assets are positive

Example: $r_i = \beta_i r_m + \sigma_i \epsilon_i$

=> $V = \beta \beta^T + \Sigma$

where r_m and ϵ_i are normalized; just need some assumption on market price of risk and maximum correlation to market

Updating to Posterior

Suppose view is given by r^{view} *and* V^{view}

Given confidence ($0 \leq \alpha \leq 1$) in view

$(r^{post}, V^{post}) = (1-\alpha)\,(r^{prior}, V^{prior}) + \alpha(r^{view}, V^{view})$

Solve with (r^{post}, V^{post}) (with caution that it may not be market consistent)

Alternatives: Chevrier (MCMC enforcing non-negative weight solutions)

Mix optimum from view and CAPM (LeDoit-Wolf)

Mix of views (like batch means)

Further Alternatives

Robust estimation (DeMiguel, Nogales)

- Remove outliers from estimates
- Solve with estimates

- Simple rules (DeMiguel, Garlappi,Uppal)
  - Just place *1/n* in each asset
  - Results: Better Sharpe ratio and lower turnover than any estimation procedure attempted
- So, is naïve diversification the best?

Some Results

Monthly data sets from MSCI and Ken French’s website as in DeMiguel, et al.

Comparisons:

- 1/N
- Moving window (120 months) estimate
- Full history estimate
- GARCH estimates

Alternative sub-strategies:

- Weight on basic CAPM prior (non-data information)
- No-short-sale constraint
- Robust optimization

Sharpe Ratios

Strategy	Weight on Prior	Industry	International	FF3	FFPortfolios+1	FFPortfolios+4
1/N	0	0.137	0.092	0.235	0.164	0.176
MV (uncon.)	0	0.213	0.160	0.278	0.761	1.764
MV (no-short)	0	0.173	0.111	0.278	0.267	0.368
MovingWindow (uncon.)	0	-0.001	-0.070	0.204	0.207	**1.554**
MovingWindow (no-short)	1	0.071	0.086	0.137	0.247	0.254
	0.5	0.078	0.093	0.171	0.253	0.291
	0	0.097	0.098	0.229	0.254	0.344
MovingWindow (Robust, uncon.)	0	0.111	0.074	0.105	0.256	0.312
MovingWindow (robust, no-short))	0	0.102	0.060	0.105	0.244	0.292
FullHistory (no-short)	1	0.102	0.086	0.210	0.230	0.247
	0.5	0.100	0.076	0.226	0.237	0.320
	0	0.105	0.075	**0.242**	0.239	0.342
GARCH (no short)	1	0.167	0.108	0.158	0.239	0.238
	0.5	0.167	0.110	0.183	0.249	0.284
	0	**0.177**	**0.121**	0.241	**0.259**	0.347
GARCH (robust, uncon.)	0	0.158	0.102	0.171	0.245	0.303
GARCH (robust, no-short)	0	0.032	0.023	0.029	0.228	0.203

Implications

Approach	Conver-gence	Better with *n*	Solution character	C.I. for all *N*	Efficient
Large sample	Y	N	M	N	M
Batch	Y	Y	Y	N	M
Robust	N	N	N	M	Y
Bayesian/Non-data Info.	Y	M	M	M	M
Naïve	N	M	Y	N	Y

Additional Issues

Non-normal distributions (Chavez/Birge (2011)):

- Mean-variance may be far from optimizing utility
- For exponential utility, can use generalized hyperbolic distributions – closed form for some examples
- Mean-variance can be close (but only if the risk-aversion parameter is chose optimally)

Additional examples:

- Non-linear functions of Gaussian distributions
- Can use polynomial approximations and higher moments to obtain optimal solutions for these non-normal cases

Summary Observations

- Convergence to asymptotic behavior may be much slower with optimization and different uncertainty forms than simple estimation
- Dimension has more effect with greater uncertainty
- Use of optimization in batches can improve estimates especially with potentially violated constraints and symmetric feasible regions
- Best MV portfolio results using GARCH-type estimates

Additional Questions

- How does the batch sample continue to improve with dimension and what are the effects of dimension in general?
- Are more general confidence interval estimates available?
- How do these approaches perform with other techniques to enhance convergence?
- What are the combined effects with estimation, non-stationarity, and non-normal distributions?

Thank you!

Can This Help with Variance Example?

- Suppose we have 10,000 assets
- Now, we need ~50,000,000 correlations to construct the variance-covariance matrix
- Problem: Analysis all assumed independence
  - If independent, then have positive definiteness problem again
  - If a single time series:
    - Observations are not independent
    - Limited number of degrees of freedom
    - Cannot estimate everything with any accuracy
- What to do?

Observations on Portfolios

- Batch approach improves when constraints can bind the sample solutions
- The batch improvement is significant when constraints are relatively tight (but still more than 3 standard deviations from optimum)
- Batch can improve without constraints (but not so much in low dimensions ~10)

Effects of Dimension

• Increase dimension from 10 to 20:

Range: $X=[0,1]^{20}$

- 982/1000 better with batching (cf. 1000/1000)
- Average improvement: 15.4% in objective (cf. 19.3%)

$X=[-1,2]^{20}$

- 840/1000 better with batching (cf. 638/1000)
- Average improvement: 8.5% in objective (cf. 3.0%)

$X=[-5,10]^{20}$

- 39/1000 better with batching (cf. 231/1000)
- Average improvement: -22% in objective (cf. -8%)

Observation: there is a "sweet spot" in constraints where batching is especially effective

Observations on Portfolios

- Batch approach improves when constraints can bind the sample solutions
- The batch improvement is significant when constraints are relatively tight (but still more than 3 standard deviations from optimum)
- Batch can improve without constraints (but not so much here)
- Batching favored by medium constraint level

Towards a General Result

In cases of symmetric bias (e.g., small example and portfolio optimization), x^{ν} may be symmetrically distributed around x^* (even though the asymptotic distribution is not obtained)

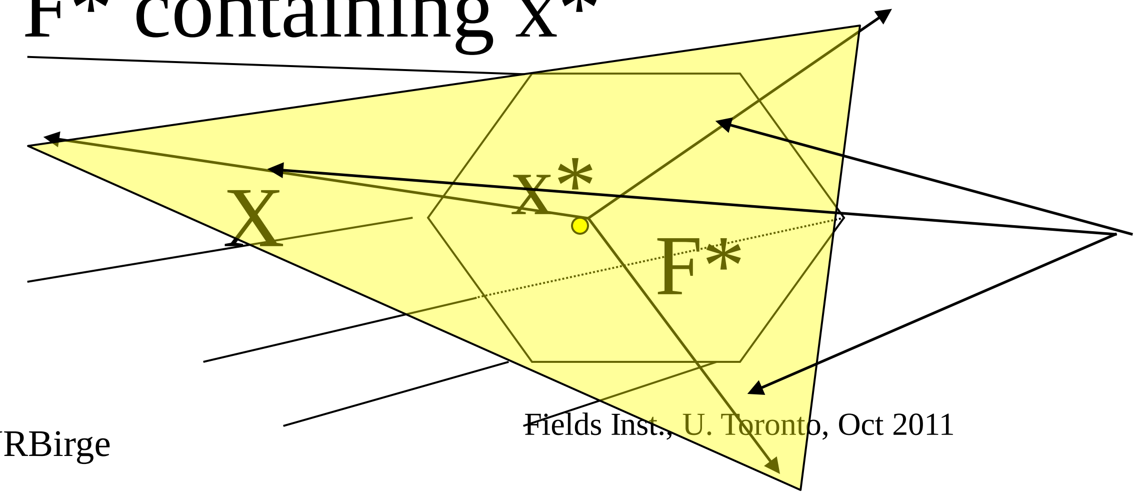
Partitioning into k independent sub-samples, $x^{\nu,k}$ may be an unbiased estimate of x^* (with error that can improve in dimension as in the example)

Requirements:

- distribution of x^{ν} for small sample sizes
- symmetry in the case of tail-risk measures may require different formulations

General Implications?

- How to put the batch results in terms of universal bounds?
- View: consider errors distributed throughout X and decompose by cone support in face F* containing x*



Assumptions

- Under mild conditions, x* is randomly distributed in F*
- Assume bias is known (or bounded)

$$b_{\nu/K} = \|E[x^{\nu/K} - x^*]\| \qquad O((\nu/K)^{-\frac{1}{2n}}$$

under certain regularity conditions (e.g., Roemisch and Schulz (1991))

- Worst error in any direction is g/n.

General Result

- Under these conditions,

$$P(\|\bar{u}^{\nu,K}\| \geq b_{\nu/K} + \frac{aM((N+1)g(N-g))^{1/2}}{K^{1/2}N}) \leq \frac{1}{a^2+1}.$$

- So, if unbiased, $a=K^{1/4}$,

Additional Considerations

- Performance is also U-shaped in batch size (when batching is improving)
- Some indication in analytical results for simple problems
- Difficult to assess optimal batch sizes in practice without experimentation