Estimation and Optimization of Portfolio Allocations

John Birge
University of Chicago
Booth School of Business

© JRBirge
INFORMS Charlotte, Nov. 2011
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
Example: Financial Portfolio Optimization

Quadratic program (Markowitz Portfolio):

find investments $x=(x(1),\ldots,x(n))$ to

$$\text{min } x^T Q x$$

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:

<table>
<thead>
<tr>
<th>DomCommon</th>
<th>SmallCap</th>
<th>IntCommon</th>
<th>EmerMarkets</th>
<th>AbsoluteRetu</th>
<th>VentCap</th>
<th>RealEst</th>
<th>Oil and Gas</th>
<th>Commodities</th>
<th>FixedIncome</th>
<th>IntFixedInc</th>
<th>Cash</th>
</tr>
</thead>
<tbody>
<tr>
<td>DomCommon</td>
<td>1</td>
<td>0.79</td>
<td>0.58</td>
<td>0.56</td>
<td>0.6</td>
<td>0.44</td>
<td>0.25</td>
<td>0.01</td>
<td>-0.3</td>
<td>0.43</td>
<td>0.2</td>
</tr>
<tr>
<td>SmallCap</td>
<td>0.79</td>
<td>1</td>
<td>0.48</td>
<td>0.61</td>
<td>0.65</td>
<td>0.56</td>
<td>0.24</td>
<td>0.01</td>
<td>-0.05</td>
<td>0.31</td>
<td>0.1</td>
</tr>
<tr>
<td>IntCommon</td>
<td>0.58</td>
<td>0.48</td>
<td>1</td>
<td>0.37</td>
<td>0.45</td>
<td>0.25</td>
<td>0.38</td>
<td>-0.04</td>
<td>-0.17</td>
<td>0.35</td>
<td>0.55</td>
</tr>
<tr>
<td>EmerMarkets</td>
<td>0.56</td>
<td>0.61</td>
<td>0.37</td>
<td>1</td>
<td>0.3</td>
<td>0.07</td>
<td>-0.19</td>
<td>-0.07</td>
<td>-0.07</td>
<td>0.1</td>
<td>0.04</td>
</tr>
<tr>
<td>AbsoluteRetu</td>
<td>0.6</td>
<td>0.65</td>
<td>0.45</td>
<td>0.3</td>
<td>1</td>
<td>0.35</td>
<td>0.2</td>
<td>-0.2</td>
<td>0.11</td>
<td>0.35</td>
<td>0.25</td>
</tr>
<tr>
<td>VentCap</td>
<td>0.44</td>
<td>0.56</td>
<td>0.25</td>
<td>0.3</td>
<td>0.35</td>
<td>1</td>
<td>0.21</td>
<td>-0.02</td>
<td>-0.18</td>
<td>0.19</td>
<td>0.15</td>
</tr>
<tr>
<td>RealEst</td>
<td>0.25</td>
<td>0.24</td>
<td>0.38</td>
<td>0.07</td>
<td>0.2</td>
<td>0.21</td>
<td>1</td>
<td>0.08</td>
<td>-0.53</td>
<td>0.15</td>
<td>0.2</td>
</tr>
<tr>
<td>Oil and Gas</td>
<td>0.01</td>
<td>0.01</td>
<td>-0.04</td>
<td>-0.19</td>
<td>-0.2</td>
<td>-0.02</td>
<td>0.08</td>
<td>1</td>
<td>0.54</td>
<td>-0.18</td>
<td>-0.3</td>
</tr>
<tr>
<td>Commodities</td>
<td>-0.3</td>
<td>-0.05</td>
<td>-0.17</td>
<td>-0.07</td>
<td>0.11</td>
<td>-0.18</td>
<td>-0.53</td>
<td>0.54</td>
<td>1</td>
<td>-0.3</td>
<td>-0.08</td>
</tr>
<tr>
<td>FixedIncome</td>
<td>0.43</td>
<td>0.31</td>
<td>0.35</td>
<td>-0.07</td>
<td>0.35</td>
<td>0.19</td>
<td>-0.18</td>
<td>-0.3</td>
<td>-0.3</td>
<td>1</td>
<td>0.55</td>
</tr>
<tr>
<td>IntFixedInc</td>
<td>0.2</td>
<td>0.1</td>
<td>0.55</td>
<td>0.1</td>
<td>0.25</td>
<td>0.15</td>
<td>0.2</td>
<td>-0.3</td>
<td>-0.08</td>
<td>0.55</td>
<td>1</td>
</tr>
<tr>
<td>Cash</td>
<td>0.27</td>
<td>0.08</td>
<td>0.23</td>
<td>0.04</td>
<td>0.45</td>
<td>0.14</td>
<td>0.37</td>
<td>-0.07</td>
<td>-0.13</td>
<td>0.67</td>
<td>0.1</td>
</tr>
</tbody>
</table>
### Results from Optimization

<table>
<thead>
<tr>
<th>Category</th>
<th>Amount to Invest</th>
</tr>
</thead>
<tbody>
<tr>
<td>DomCommon</td>
<td>-54079107483</td>
</tr>
<tr>
<td>SmallCap</td>
<td>-17314640180</td>
</tr>
<tr>
<td>InteCommon</td>
<td>-7098209713</td>
</tr>
<tr>
<td>EmerMarkets</td>
<td>21285151081</td>
</tr>
<tr>
<td>AbsoluteReturn</td>
<td>65911278496</td>
</tr>
<tr>
<td>VentCap</td>
<td>3346118938</td>
</tr>
<tr>
<td>RealEst</td>
<td>-68300117028</td>
</tr>
<tr>
<td>Oil and Gas</td>
<td>66227880617</td>
</tr>
<tr>
<td>Commodities</td>
<td>-1.04264E+11</td>
</tr>
<tr>
<td>FixedIncome</td>
<td>-72656761796</td>
</tr>
<tr>
<td>IntFixedInc</td>
<td>1.17885E+11</td>
</tr>
<tr>
<td>Cash</td>
<td>49057530702</td>
</tr>
</tbody>
</table>

### What happened here?

- **Return**: 0.0999999487
- **Variance**: -1.64591E+19
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 \( nru \)
General Problem

- 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):

\[
\text{Min}_{x \in X} \ E_\xi [f(x, \xi)] \quad (P)
\]

• Suppose the only information for \( \xi \) is through observations: \( \xi^1, \ldots, \xi^v \)

• Typical empirical case:

\[
\text{Min}_{x \in X} \ (1/ v) \sum_{i=1}^{v} f(x, \xi^i)
\]

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

• What are the best ways to use those observations?
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)
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 $\nu$ (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?
Portfolio Context

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?
Alternative Responses

Very large samples (not enough data)
Batching or sub-sample optimization
  Close to “re-sampled portfolio” ideas (Michaud)
  (Note: convergence issues)
Robust optimization
Bayesian updating
Robust estimation
Estimation with non-data information
Simple rules
Sub-sample or Batch

• Use of sub-samples or batch means (e.g., Mak, Morton, Wood (99))

• Suppose that we divide the \( \nu \) samples into \( k \) batches of \( \nu/k \) each, let \( \xi_{i}^{\nu} \) be the mean of batch \( i = 1, \ldots, k \), then solve with \( \xi_{i}^{\nu} \) to obtain \( x_{i}^{\nu} \)

• Let \( x_{i}^{\nu,k} = (1/k) \sum_{i=1}^{k} x_{i}^{\nu} \)

• 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 $\mu^\nu$ and sample variance $\Sigma^\nu$ with n samples

$$x^\nu = (\Sigma^\nu)^{-1}(\mu^\nu)(\nu-n-2)/\nu$$

is an unbiased estimator of $x^*$ (for unconstrained case with risk-free asset)

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

$$(n)(\nu-n-2)/\nu^2 + (\mu \Sigma^{-1} \mu)(\nu-n-2)/\nu$$

Note: dependence on $n$;

With batches:

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

(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?

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

• Now, decreased dependence on \( n \)
Observe: more improvement as $\nu \uparrow$ (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, \(\xi\) and \(\gamma\) are random
Suppose \(\xi_j \sim \text{N}(0,1)\); \(\gamma \sim \text{N}(1,1)\)

• Unconstrained solution:
Error in solution in 2-norm is \(\chi^2\) 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:

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

and \((\nu-n-2)/\nu\) 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 $\Sigma = 0.05 \times I$, where $I$ is an identity matrix. We present the results from 1000 simulation runs for three different sets, $X$, corresponding to increasing ranges on $x$: $[0, 1]^T$, $[-1, 2]^T$, and $[-5, 10]^T$. The results are compared relative to the optimal solution $x^* = 0.4e$ in terms of $\|x' - x^*\|/\|x^*\|$ and optimal objective value $z^* = -\bar{r}^T x^* + \frac{1}{2} x^T \Sigma x^* = -0.04$ in terms of $(\bar{r}^T x' + \frac{1}{2} (x')^T \Sigma x' - z^*)/(-z^*)$.

Observe: histograms of relative errors in solutions and losses in objective
\( X = [0, 1]^{10} \)

 Relatives differences:

 Batch better: 1000/1000


 Avg. Obj. Diff.: -19%

 © JRBirge
Dimension Effect: \( X = [-1, 2] \)

Relative Distance from Optimum

\[
\begin{array}{c}
\text{Number of Observations} \\
\begin{array}{c}
\text{Relative Distance } \| e_n^K \|_2, \| e_n \|_2 \\
\end{array}
\end{array}
\]

\( n = 10 \)

\( n = 20 \)

© JRBirge

INFORMS Charlotte, Nov. 2011
Distance Effect: \( X = [-5, 10] \)

Relative Distance from Optimum

© JRBirge

INFORMS Charlotte, Nov. 2011
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)
Robust Optimization

Idea: Suppose an uncertainty set around the estimated data
Optimize over the worst case in the uncertainty set
Example: \((r, V) \in \mathcal{R} \times \mathcal{W}\)

\[
\min \left( \max_{(r,V) \in \mathcal{R} \times \mathcal{W}} x^T V x \right) \\
\text{s.t. } r^T x \geq r^*, \quad e^T x = 1 \quad (x \geq 0)
\]
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 covariances (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 \( \epsilon_i \) are normalized; just need some assumption on market price of risk and maximum correlation to market

© JRBirge
Updating to Posterior

Suppose view is given by $r^{\text{view}}$ and $V^{\text{view}}$

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

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

Solve with $(r^{\text{post}}, V^{\text{post}})$ (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
<table>
<thead>
<tr>
<th>Strategy</th>
<th>Weight on Prior</th>
<th>Industry</th>
<th>International</th>
<th>FF3</th>
<th>FFPorfolios+ 1</th>
<th>FFPorfolios+ 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/N</td>
<td>0</td>
<td>0.137</td>
<td>0.092</td>
<td>0.235</td>
<td>0.164</td>
<td>0.176</td>
</tr>
<tr>
<td>MV (uncon.)</td>
<td>0</td>
<td>0.213</td>
<td>0.160</td>
<td>0.278</td>
<td>0.761</td>
<td>1.764</td>
</tr>
<tr>
<td>MV (no-short)</td>
<td>0</td>
<td>0.173</td>
<td>0.111</td>
<td>0.278</td>
<td>0.267</td>
<td>0.368</td>
</tr>
<tr>
<td>MovingWindow (uncon.)</td>
<td>0</td>
<td>-0.001</td>
<td>-0.070</td>
<td>0.204</td>
<td>0.207</td>
<td>1.554</td>
</tr>
<tr>
<td>MovingWindow (no-short)</td>
<td>1</td>
<td>0.071</td>
<td>0.086</td>
<td>0.137</td>
<td>0.247</td>
<td>0.254</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.078</td>
<td>0.093</td>
<td>0.171</td>
<td>0.253</td>
<td>0.291</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.097</td>
<td>0.098</td>
<td>0.229</td>
<td>0.254</td>
<td>0.344</td>
</tr>
<tr>
<td>MovingWindow (Robust, uncon.)</td>
<td>0</td>
<td>0.111</td>
<td>0.074</td>
<td>0.105</td>
<td>0.256</td>
<td>0.312</td>
</tr>
<tr>
<td>MovingWindow (robust, no-short)</td>
<td>0</td>
<td>0.102</td>
<td>0.060</td>
<td>0.105</td>
<td>0.244</td>
<td>0.292</td>
</tr>
<tr>
<td>FullHistory (no-short)</td>
<td>1</td>
<td>0.102</td>
<td>0.086</td>
<td>0.210</td>
<td>0.230</td>
<td>0.247</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.100</td>
<td>0.076</td>
<td>0.226</td>
<td>0.237</td>
<td>0.320</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.105</td>
<td>0.075</td>
<td>0.242</td>
<td>0.239</td>
<td>0.342</td>
</tr>
<tr>
<td>GARCH (no short)</td>
<td>1</td>
<td>0.167</td>
<td>0.108</td>
<td>0.158</td>
<td>0.239</td>
<td>0.238</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.167</td>
<td>0.110</td>
<td>0.183</td>
<td>0.249</td>
<td>0.284</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.177</td>
<td>0.121</td>
<td>0.241</td>
<td>0.259</td>
<td>0.347</td>
</tr>
<tr>
<td>GARCH (robust, uncon.)</td>
<td>0</td>
<td>0.158</td>
<td>0.102</td>
<td>0.171</td>
<td>0.245</td>
<td>0.303</td>
</tr>
<tr>
<td>GARCH (robust, no-short)</td>
<td>0</td>
<td>0.032</td>
<td>0.023</td>
<td>0.029</td>
<td>0.228</td>
<td>0.203</td>
</tr>
</tbody>
</table>
## What to Do?

<table>
<thead>
<tr>
<th>Approach</th>
<th>Convergence</th>
<th>Better with $n$</th>
<th>Solution character</th>
<th>C.I. for all $N$</th>
<th>Efficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Large sample</td>
<td>Y</td>
<td>N</td>
<td>M</td>
<td>N</td>
<td>M</td>
</tr>
<tr>
<td>Batch</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>M</td>
</tr>
<tr>
<td>Robust</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>M</td>
<td>Y</td>
</tr>
<tr>
<td>Bayesian/Non-data Info.</td>
<td>Y</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>Naïve</td>
<td>N</td>
<td>M</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
</tr>
</tbody>
</table>
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!