Portfolio Optimization: A Brief Tutorial

John Birge
University of Chicago
Booth School of Business
Overview

- Portfolio optimization involves:
  - Modeling
  - Optimization
  - Estimation
  - Dynamics

- Key issues:
  - Representing utility (or risk and reward)
  - Choosing distribution classes (and parameters)
  - Solving the resulting problems
  - Implementing solutions over time with non-stationary processes, transaction costs, taxes, and uncertain future regulations
Outline

• Introduction
• Modeling
• Estimation
• Solutions
• Implementation
• Dynamics
• Conclusion
Basic Problem Setup

Choose an allocation $x \in \mathbb{R}^n$ across $n$ assets (classes) to maximize expected utility at time $T$: $E[u_T(x)]$

Note: $x$ may be a process $x_t$ but not considering consumption and liabilities.
Utility Forms/Risk Metrics

• CARA/exponential: with wealth:
  \[ w_T = \rho_T x_T, \quad u_T(x_T) = 1 - e^{-\gamma w_T} \]

• CRRA: \[ u_T(x_T) = w_T^{\gamma} / \gamma \]

• Risk-reward:
  Max Reward – Risk or Reward/Risk
  Max Reward s.t. Risk constraint
  Min Risk s.t. Reward constraint
Forms of Reward/Risk

- Reward: expected return: $E(r)^T x$
- Risk:
  - Variance or standard deviation of return
    $$x^T V x \text{ or } (x^T V x)^{1/2}$$
  - Semi-deviation
  - Value-at-Risk
  - Conditional Value-at-Risk
Mean-Variance (Markowitz) Model

Model:

Choose $x \in \mathbb{R}^n$ to:

- minimize $x^T V x$
- s.t. $E(r)^T x \geq r_0$, $e^T x = 1$, $x \in X$.

Issues:

How to find $V$ and $E(r)$?

Naive response: Use the sample variance and mean.
Estimation Issues

Quadratic program: \( \text{min } x^T V x \)

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

where \( V \) and \( E(r) \) are estimated from historical data.

Results from sample estimates:

Example: correlations from University of Michigan CIO:

<table>
<thead>
<tr>
<th>DomCommon</th>
<th>SmallCap</th>
<th>InteCommon</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>InteCommon</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>0.14</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>0.15</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>1</td>
<td>0.55</td>
<td>0.67</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 $nru$
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)
Incorporate estimation error in optimization
Batch or sub-sample or “re-sampled portfolio” (Michaud)
Robust optimization
Bayesian updating
Factor models
Robust estimation
Simple rules
Sample Estimation and Optimization

Use of sample estimates (normal and stationary):
(See, e.g., Kan and Zhou (2007).)
Sample mean return with $N$ samples: $\hat{r}$ is $N(\bar{r}, V/N)$
Sample variance: $\hat{V}$ is $W_n(N-1, V)/N$
(Wishart with $N-1$ dof)
$E(\hat{V}^{-1} \hat{\mu}) = \frac{N}{N-n-2} V^{-1} \bar{r}$.
→ bias in sample estimator for optimization

→ use multiplier of $\frac{N-n-2}{N}$ in objective
for tangency (maximum Sharpe ratio)
i.e., shrinkage estimation of variance.

© JRBirge
INFORMS Minneapolis, Oct. 2013
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 $\xi^\nu_i$ be the mean of batch $i=1,\ldots,k$, then solve with $\xi^\nu_i$ to obtain $x^\nu_i$
• 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 $\mu^\nu$ 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
$X = [0, 1]^{10}$

Relatives differences:
Batch better: 1000/1000
Avg. Obj. Diff.: -19%

Solution

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

$n=10$  
$n=20$

© JRBirge

INFORMS Minneapolis, Oct. 2013
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

(e.g., Goldfarb/Iyengar, Ceria/Stubbs, Tütüncü/Koenig, Fabozzi, et al.
Distributional: Delage/Ye)

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

\[
\begin{align*}
\operatorname{Min} \ (\operatorname{Max}_{(r,V) \in \mathcal{R} \times \mathcal{W}} x^T V x) \\
\text{s.t. } r^T x \geq r^*, \ e^T x = 1 \ (x \geq 0))
\end{align*}
\]
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: $r_i = \beta_i r_m + \sigma_i \epsilon_i$

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

where $r_m$ and $\epsilon_i$ are normalized; just need some assumption on market price of risk and maximum correlation to market
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)
Factor Models

Basic idea:
Estimate each return as a function of a set of factors:

\[ r = (\alpha) + Bf + \epsilon, \]

Estimation:
Now, just estimate \( \beta_i \) for each asset \( i \) (e.g., from regression).
Factors: Fama-French (HML, MKT, SMB), Momentum, Region, Industry, etc.

Issues:
Still: much to estimate.
Missing factors? (Factor alignment, e.g., Saxeena and Stubbs (2011))
Nonstationarity? (GARCH - smoothing over time)
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>FFPortfolios+ 1</th>
<th>FFPortfolios+ 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>

© JR Birge
### 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-Bedoya/Birge):
- 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 approaches:
- Non-linear functions of Gaussian distributions
- Can use polynomial approximations and higher moments to obtain optimal solutions for these non-normal cases
Transaction Costs/Taxes and Dynamics

Transaction costs:
Each trade has some impact (e.g., bid-ask spread plus commission). Large trades may have long-term impacts.

Taxes:
Taxes depend on the basis and vintage of an asset and involve alternative selling strategies (LIFO, FIFO, lowest/highest price).
Why Model Dynamically?

Three potential reasons:
  Market timing
  Reduce transaction costs (taxes) over time
  Maximize wealth-dependent objectives

Example
  Suppose major goal is $100MM to pay pension liability in 2 years
  Start with $82MM; Invest in stock (annual vol=18.75%, annual exp. Return=7.75%); bond (Treasury, annual vol=0; return=3%)
  Can we meet liability (without corporate contribution)?
  How likely is a surplus?
Alternatives

Markowitz (mean-variance) – Fixed Mix

Pick a portfolio on the efficient frontier
Maintain the ratio of stock to bonds to minimize expected shortfall

Buy-and-hold (Minimize expected loss)
Invest in stock and bonds and hold for 2 years

Dynamic (stochastic program)
Allow trading before 2 years that might change the mix of stock and bonds
Efficient Frontier

Some mix of risk-less and risky asset

For 2-year returns:
Best Fixed Mix and Buy-and-Hold

Fixed Mix: 27% in stock
Meet the liability 25% of time (with binomial model)

Buy-and-Hold: 25% in stock
Meet the liability 25% of time
Best Dynamic Strategy

Start with 57% in stock

If stocks go up in 1 year, shift to 0% in bond
If stocks go down in 1 year, shift to 91% in stock
Meet the liability 75% of time
Advantages of Dynamic Mix

Able to lock in gains
Take on more risk when necessary to meet targets
Respond to individual utility that depends on level of wealth
Approaches for Dynamic Portfolios

Static extensions

- Can re-solve (but hard to maintain consistent objective)
- Solutions can vary greatly
- Transaction costs difficult to include

Dynamic programming policies

- Approximation
- Restricted policies (optimal – feasible?)
- Portfolio replication (duration match)

General methods (stochastic programs)

- Can include wide variety
- Computational (and modeling) challenges
Basic Model with Transaction Costs

• Basic setup:

Find \( x(t), b(t), s(t) \) to maximize \( E(u(x(T))) \) subject to \( x(0) \):

\[
e^T x^+(t) = e^T x(t) - \tau^T b(t) - \tau^T s(t),
\]

\[
e^T (b(t) + s(t)) = 0,
\]

\[
x^+(t) + (I + \text{diag}(\tau)) s(t) - (I - \text{diag}(\tau)) b(t) = x(t),
\]

where \( \tau \) represents transaction costs and \( x(0) \) gives initial conditions and, without control, \( x(t) \) follows geometric Brownian motion

\[
dx(t) = x(t)(\mu(t) + \Sigma(t)^{1/2}dW(t))
\]

where \( W(t) \) represents \( n \) independent Brownian motions.
Continuous-Time Results


Results: No trading in a region $H$; boundary at some distance from optimal no-transaction-cost point (for CRRA utility:

$$x^*=(1/g)\Sigma^1(\mu-r), \text{ Merton line}$$
General Result

$x_1(t)$

Merton line

No-trade region

Time $T$

© JRBirge

INFORMS Phoenix, October 2012
Equivalence in Discrete Time

General observation: The continuous time solution is (approximately) equal to a discrete-time problem with a fixed boundary.

$x_1(t)$

Merton line

No-trade region

Boundary here: same as for one period to $T^*$.  

$T^*$

Time

$T$

© JRBirge  

INFORMS Phoenix, October 2012
Discrete Time (Single Period) Problem

Find $x, b, s$ to minimize $E[u(x(t + T^*))]$ s.t.

$$e^T x(t) = e^T x_0(t) - \tau^T b(t) - \tau^T s(t),$$

$$e^T (b(t) + s(t)) = 0,$$

$$x(t) + (I + \text{diag}(\tau)) s(t) - (I - \text{diag}(\tau)) b(t) = x_0,$$

where $e^T x_0 = w(t)$.

Challenge: How to find $T^*$?
Effective Result in Terms of Average Number of Re-balances

Observation: \( T^* \) is approximately the average time between re-balances or \( 1/T^* \) is approximately the expected number of re-balances in a single period.

- Can normalize to a single period and use \( \pi/T^* \) for transaction cost.

- (Note: can learn \( T^* \) along with \( \mu, \Sigma \))
Dynamic Programming Approach

State: $x_t$, corresponding to positions in each asset (and possibly price, economic, other factors)

Value function: $V_t(x_t)$

Actions: $u_t$

Possible events $s_t$, probability $p_{st}$

Find:

$$V_t(x_t) = \max \left\{ -c_t u_t + \sum_{st} p_{st} V_{t+1}(x_{t+1}(x_t, u_t, s_t)) \right\}$$

Advantages: general, dynamic, can limit types of policies

Disadvantages: Dimensionality, approximation of $V$ at some point needed, limited policy set may be needed, accuracy hard to judge

General Form in Discrete Time

Find $x=(x_1,x_2,...,x_T)$ and $p$ (allows for “robust formulation”) to

\[
\min_{E_p} \left[ \sum_{t=1}^{T} f_t(x_t, x_{t+1}, p) \right]
\]

s.t. $x_t \in X_t$, $x_t$ nonanticipative, $p \in P$ (distribution class)

$P[ h_t(x_t, x_{t+1}, p_t) \leq 0 ] \geq a$ (chance constraint)

General Approaches:
Simplify distribution (e.g., sample) and form a mathematical program:

• Solve step-by-step (dynamic program)
• Solve as single large-scale optimization problem

Use iterative procedure of sampling and optimization steps
What about Continuous Time?

Sometimes very useful to develop overall structure of value function

May help to identify a policy that can be explored in discrete time (e.g., portfolio no-trade region)

Analysis can become complex for multiple state variables

Possible bounding results for discrete approximations (e.g., FEM approach)
Restricted Policy Approaches:

1. Fixed proportions
2. Fixed function of factors/state variables
3. Contingent functions

ADP Approaches:
Approximate value function $V_t(x_t)$ by a combination of basis functions:

$$V_t(x_t) = \sum_{i} \lambda_i \phi_i(x_t)$$

and optimize over weights $\lambda$. 

© JRBirge INFORMS Minneapolis, Oct. 2013
Large-Scale Optimization

Basic Framework: Stochastic Programming

Model Formulation:

\[
\text{max } \sum_{\sigma} p(\sigma) \left( U(W(\sigma, T)) \right)
\]

s.t. (for all \( \sigma \)):

\[
\sum_k x(k, 1, \sigma) = W(0) \quad \text{(initial)}
\]

\[
\sum_k r(k, t-1, \sigma) \cdot x(k, t-1, \sigma) - \sum_k x(k, t, \sigma) = 0, \text{ all } t > 1;
\]

\[
\sum_k r(k, T-1, \sigma) \cdot x(k, T-1, \sigma) - W(\sigma, T) = 0, \text{ (final)};
\]

\[
x(k, t, \sigma) \geq 0, \text{ all } k, t;
\]

Nonanticipativity:

\[
x(k, t, \sigma') - x(k, t, \sigma) = 0 \text{ if } \sigma', \sigma \in S_i, \text{ for all}
\]

Advantages:

General model, can handle transaction costs, include tax lots, etc.

Disadvantages: Size of model, insight
Simplified Finite Sample Model

Assume $p$ is fixed and random variables represented by sample $\xi_t$ for $t=1,2,\ldots,T$, $i=1,\ldots,N_t$ with probabilities $p^i_t$, $a(i)$ an ancestor of $i$, then model becomes (no chance constraints):

$$\begin{align*}
\text{minimize} & \quad \Sigma_{t=1}^T \Sigma_{i=1}^{N_t} p^i_t f_t(x^{a(i)}_t, x^i_{t+1}, \xi_t) \\
\text{s.t.} & \quad x^i_t \in X^i_t
\end{align*}$$

Observations?

- Problems for different $i$ are similar – solving one may help to solve others
- Problems may decompose across $i$ and across $t$ yielding
  - smaller problems (that may scale linearly in size)
  - opportunities for parallel computation.
Model Consistency

Price dynamics may have inherent arbitrage

Example: model includes option in formulation that is not the present value of future values in model (in risk-neutral prob.)

Does not include all market securities available

Policy inconsistency

May not have inherent arbitrage but inclusion of market instrument may create arbitrage opportunity

Skews results to follow policy constraints

Lack of extreme cases

Limited set of policies may avoid extreme cases that drive solutions
Objective Consistency

Examples with non-coherent objectives

Value-at-Risk
Probability of beating benchmark

Coherent measures of risk

Can lead to piecewise linear utility function forms
Expected shortfall, downside risk, or conditional value-at-risk (Uryasiev and Rockafellar)
Model and Method Difficulties

Model Difficulties

- Arbitrage in tree
- Loss of extreme cases
- Inconsistent utilities

Method Difficulties

- Deterministic incapable on large problems
- Stochastic methods have bias difficulties
  - Particularly for decomposition methods
  - Discrete time approximations
- Stopping rules and time hard to judge
Resolving Inconsistencies

Objective: Coherent measures (& good estimation)

Model resolutions

  Construction of no-arbitrage trees (e.g., Klaassen)
  Extreme cases (Generalized moment problems and fitting with existing price observations)

Method resolutions

  Use structure for consistent bound estimates
  Decompose for efficient solution
Abridged Nested Decomposition

Incorporates sampling into the general framework of Nested Decomposition
Assumes relatively complete recourse and serial independence
Samples both the sub-problems to solve and the solutions to continue from in the forward pass through sample-path tree
Lagrangian-based Approaches

General idea:

Relax nonanticipativity (or perhaps other constraints)
Place in objective

Separable problems

\[
\begin{align*}
\text{MIN} & \quad E \left[ \sum_{t=1}^{T} f_t(x_t, x_{t+1}) \right] \\
\text{s.t.} & \quad x_t \in X_t \\
& \quad x_t \text{ nonanticipative}
\end{align*}
\]

\[
\begin{align*}
\text{MIN} & \quad E \left[ \sum_{t=1}^{T} f_t(x_t, x_{t+1}) \right] \\
& \quad x_t \in X_t \\
& \quad + E[w_x] + r/2||x-x||^2
\end{align*}
\]

Update: \( w_t \); Project: \( x \) into \( N \) - nonanticipative space as \( x \)

Convergence: Convex problems - Progressive Hedging Alg. (Rockafellar and Wets)
Advantage: Maintain problem structure (e.g., network)
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!