Statistical and Algorithm Aspects of Optimal Portfolios

Howard Shek

ICME, Stanford University

17 September, 2010
"A good portfolio is more than a long list of good stocks and bonds. It is a balanced whole, providing the investor with protections and opportunities with respect to a wide range of contingencies."

- Harry Markowitz in his 1959 book "Portfolio Selection: Efficient Diversification of Investments"
Cardinally Constrained Portfolio Optimization

*NP-Hard* combinatorial problem

\[
f^* = \min_x c^\top x + \frac{1}{2} x^\top H x \\
s.t. \quad A x \geq b \\
\sum_i 1\{x_i \neq 0\} = K
\]

where \( x \) are the weights, \( 1\{x_i \neq 0\} \) is the indicator function for non-zero weight assets, and \( K \) is the *cardinality* constraint. NP-hard problems require algorithms of exponential complexity.
Outline

1. First Aspect: Covariance Estimation
   - Background
   - RealGARCH Framework, Hansen, Huang & Shek (2009)

2. Second Aspect: Expected Return Estimation
   - Current Practice
   - Marked Point Process with Limit Orderbook, Shek (2009)

3. Third Aspect: Cardinaly Constrained Optimal Portfolio
   - Current Practice
   - Local Smoothing Algorithm, Murray & Shek (2010)
High Frequency Data
Increasing influence of HFT, better access to trade and order book data

Figure: NYSE HFT: 60% in '09; 30% in '05. Barcap.
High Frequency Return Models
Covariance and expected return estimation and prediction

H: variance-covariance matrix estimation and prediction
RealGARCH univariate and multivariate framework for covariance matrix estimation and prediction. Hansen et al. (2009)

C: expected return model
Marked point process incorporating order book dynamics. Shek (2010)
Cardinally constrained optimization


Optimal execution with limit and market orders

First Aspect: Covariance Estimation

Covariance Matrix Estimation and Prediction

Make estimation and prediction for the variance-covariance matrix $H$, using high frequency information.
Absence of Noise
Latent log-price $X_t$ a Wiener process

\[ dX_t = \mu_t dt + \sigma_t dW_t \]

**Definition**

The Integrated Variance (IV) or Quadratic Variation (QV) of the process is defined to be

\[ \langle X, X \rangle_T = \int_0^T \sigma_t^2 dt \]
Absence of Noise
Variance estimator

When the sampling interval over the period $[0, T]$ is equally spaced at $\Delta$, the realized variance becomes

$$[X, X]_T = \sum_{j=1}^{T/\Delta} \left( X_{t\Delta j} - X_{t\Delta(j-1)} \right)^2$$
Presence of Market Microstructure Noise

Corruption of latent price by noise

Return process observed at the sampling times is of the form

\[ Y_{t_i} = X_{t_i} + \epsilon_{t_i} \]

- Bid-ask bounce, where the price fluctuates between the prevailing bid and ask prices;
- Discretized trading, where prices move in discrete increments.
Presence of Market Microstructure Noise

Tick by tick realized variance: \( \langle X, X \rangle^{(\text{tick})}_T \)

\[
[X, X]_T + 2nE[\varepsilon^2]
\]

object of interest due to noise

where \( n = T/\Delta \).
Presence of Market Microstructure Noise

Realized kernel estimator: $\langle X, X \rangle^{(rk)}_T$

\[
\sum_{h=-H}^{H} K \left( \frac{h}{H+1} \right) \sum_{j=|h|+1}^{T/\Delta} \left( Y_{\Delta j} - Y_{\Delta(j-1)} \right) \left( Y_{\Delta(j-h)} - Y_{\Delta(j-h-1)} \right)
\]

Parzen kernel:

\[
K(x) = \begin{cases} 
1 - 6x^2 + 6x^3 & 0 \leq x \leq 1/2 \\
2(1-x)^3 & 1/2 \leq x \leq 1 \\
0 & x > 1
\end{cases}
\]

Asymptotic consistency:

\[
\langle X, X \rangle^{(rk)}_T \xrightarrow{p} \int_0^T \sigma_s \, ds
\]
First Aspect: Covariance Estimation
Second Aspect: Expected Return Estimation
Third Aspect: Cardinaly Constrained Optimal Portfolio

Contribution

Background
RealGARCH Framework, Hansen, Huang & Shek (2009)

Framework
Univariate RealGARCH\((p,q)\)

\[
\begin{align*}
  r_t &= \sqrt{h_t} z_t, \quad \text{var} \left( r_t | F_{t-1} \right) = h_t \\
  \log h_t &= \omega + \sum_{i=1}^{p} \beta_i \log h_{t-i} + \sum_{j=1}^{q} \gamma_j \log x_{t-j} \\
  \log x_t &= \xi + \varphi \log h_t + \tau(z_t) + u_t
\end{align*}
\]

where \( z_t = r_t / \sqrt{h_t} \sim iid(0,1), \ u_t \sim iid(0, \sigma_u^2) \), and \( \tau(z) = \tau_1 z + \tau_2 (z^2 - 1) \) is called the leverage function, \( x_t \) is realized kernel estimator.

Realized Kernel Estimator, \( x_t \), is

- a better proxy for intra-day quadratic variance than \( r_t^2 \),
- linked to the conditional variance, \( h_t \), the measurement equation.
### Fitted Result for SPY

<table>
<thead>
<tr>
<th>Model</th>
<th>GARCH(1,1)</th>
<th>RealGARCH(1,1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega$</td>
<td>0.04</td>
<td>0.06</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.03</td>
<td>-</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.96</td>
<td>0.55</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>-</td>
<td>0.41</td>
</tr>
<tr>
<td>$\xi$</td>
<td>-</td>
<td>-0.18</td>
</tr>
<tr>
<td>$\phi$</td>
<td>-</td>
<td>1.04</td>
</tr>
<tr>
<td>$\sigma_u$</td>
<td>-</td>
<td>0.38</td>
</tr>
<tr>
<td>$\tau_1$</td>
<td>-</td>
<td>-0.07</td>
</tr>
<tr>
<td>$\tau_2$</td>
<td>-</td>
<td>0.07</td>
</tr>
<tr>
<td>$\ell(r,x)$</td>
<td>-</td>
<td>-2395.6</td>
</tr>
<tr>
<td>$\pi$</td>
<td>0.988</td>
<td>0.975</td>
</tr>
<tr>
<td>$\ell(r)$</td>
<td>-1752.7</td>
<td>-1712.0</td>
</tr>
</tbody>
</table>
Leverage Effect
Effect captured by leverage function $\tau(z)$

Figure: News impact curve for IBM and SPY
First Aspect: Covariance Estimation
Second Aspect: Expected Return Estimation
Third Aspect: Cardinally Constrained Optimal Portfolio
Contribution

Background
RealGARCH Framework, Hansen, Huang & Shek (2009)

Multiperiod Forecast

\[
Y_{t+h} = A^h Y_t + \sum_{j=0}^{h-1} A^j (b + \varepsilon_{t+h-j}), \quad \varepsilon_t = \begin{bmatrix} 0_{p \times 1} \\ \tau(z_t) + u_t \\ 0_{q \times 1} \end{bmatrix}
\]

\[
Y_t = \begin{bmatrix} \log h_t \\ \vdots \\ \log h_{t-p+1} \\ \log x_t \\ \vdots \\ \log x_{t-q+1} \end{bmatrix}, \quad A = \begin{pmatrix} (\beta_1, \ldots, \beta_p) & (\gamma_1, \ldots, \gamma_q) \\ (l_{p-1 \times p-1}, 0_{p-1 \times 1}) & 0_{p-1 \times q} \\ \varphi(\beta_1, \ldots, \beta_p) & \varphi(\gamma_1, \ldots, \gamma_q) \\ 0_{q-1 \times p} & (l_{q-1 \times q-1}, 0_{q-1 \times 1}) \end{pmatrix}
\]

\[
b = (\omega, 0_{p-1 \times 1}, \xi + \varphi \omega, 0_{q-1 \times 1})^\top
\]
Let $H_t = D_t R_t D_t$, then two steps

- conditional variance:

$$D^2_t = \text{diag}(\omega_i) + \text{diag}(\beta_i) \otimes D^2_{t-1} + \text{diag}(\gamma_i) \otimes X_{t-1}$$

where the diagonal matrix $D^2_t$ consists of $\log h_{1,t}, \ldots, \log h_{N,t}$, $X_t$ of $\log x_{1,t}, \ldots, \log x_{N,t}$ and

$$\log x_{i,t} = \xi_i + \varphi_i \log h_{i,t} + \tau(z_{i,t}) + u_{i,t}$$

- conditional correlation:

$$R_t = \text{diag}(Q_t)^{-1/2} Q_t \text{diag}(Q_t)^{-1/2}$$

$$Q_t = (1 - \alpha_{DCC} - \beta_{DCC}) \overline{Q} + \alpha_{DCC} \varepsilon_{t-1} \varepsilon_{t-1}^\top + \beta_{DCC} Q_{t-1}$$

where $\overline{Q} = E[\varepsilon_{t-1} \varepsilon_{t-1}^\top]$ is the unconditional correlation matrix of the standardized return.
### DCC-RealGARCH

Fitted result for four asset case: IBM, XOM, SPY and WMT

<table>
<thead>
<tr>
<th></th>
<th>IBM</th>
<th>XOM</th>
<th>SPY</th>
<th>WMT</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h_0$</td>
<td>0.76</td>
<td>1.17</td>
<td>0.28</td>
<td>2.39</td>
</tr>
<tr>
<td>$\omega$</td>
<td>-0.00</td>
<td>0.04</td>
<td>0.06</td>
<td>-0.03</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.64</td>
<td>0.59</td>
<td>0.55</td>
<td>0.66</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>0.36</td>
<td>0.30</td>
<td>0.41</td>
<td>0.30</td>
</tr>
<tr>
<td>$\xi$</td>
<td>0.01</td>
<td>-0.10</td>
<td>-0.18</td>
<td>0.12</td>
</tr>
<tr>
<td>$\varphi$</td>
<td>0.94</td>
<td>1.27</td>
<td>1.04</td>
<td>1.04</td>
</tr>
<tr>
<td>$\tau_1$</td>
<td>-0.04</td>
<td>-0.08</td>
<td>-0.07</td>
<td>-0.01</td>
</tr>
<tr>
<td>$\tau_2$</td>
<td>0.08</td>
<td>0.08</td>
<td>0.07</td>
<td>0.09</td>
</tr>
<tr>
<td>$\sigma_u$</td>
<td>0.39</td>
<td>0.38</td>
<td>0.38</td>
<td>0.40</td>
</tr>
</tbody>
</table>

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_{DCC}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_{DCC}$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table:** Fitted $\alpha_{DCC}$ and $\beta_{DCC}$ in the conditional correlation equation for DCC(1,1)-RealGARCH(1,1)
Second Aspect: Expected Return Estimation

**Expected Return Prediction**

Make estimation and prediction for the return vector $c$, using high frequency order book information.
Current Methodologies
Time, frequency and hybrid domain approach

- Three main domains of operation
  - Time. (e.g. Auto-Regressive-Integrated-Moving-Average)
  - Frequency. (e.g. Fourier Analysis)
  - Hybrid. (e.g. Wavelet Analysis)

- Wide spectrum of variation within each domain

- Input mainly historic return data sampled at equal time distance
A simple counting process for the number of trade events, characterized by arrival time of the trades, \( \{t_i\}_{i \in \{1,2,...,T\}} \), a sequence of strictly position random variable on \( (\Omega, F, P) \), such that \( t_0 = 0 \) and \( 0 < t_i \leq t_{i+1} \) for \( i \geq 1 \).

Intensity is stochastic

\[
\lambda_t = \mu + \int_{u < t} h(t-u) \, dN_u
\]

where \( \mu \) is the base intensity and \( h : \mathbb{R}^+ \rightarrow \mathbb{R}^+ \) parametrizes the self-excitation dynamics.
Hawkes Process
Univariate case

- If $\lambda_t$ satisfies the following SDE
  \[ d\lambda_t = \beta (\mu - \lambda_t) \, dt + \alpha dN_t \]
  
- The solution for $\lambda_t$ can then be written
  \[ \lambda_t = \mu + \alpha \int_0^t e^{-\beta(t-u)} dN_u \]

where we can think of $\mu$ as the long run "base" intensity, i.e. the intensity if there have been no past arrival.
Bivariate Hawkes

Bivariate Hawkes process

\[
\begin{align*}
\lambda_1(t) &= \mu_1 + \int_0^t h_{11}(t-s) \, dN_1(s) + \int_0^t h_{12}(t-s) \, dN_2(s) \\
\lambda_2(t) &= \mu_2 + \int_0^t h_{21}(t-s) \, dN_1(s) + \int_0^t h_{22}(t-s) \, dN_2(s)
\end{align*}
\]

Hawkes (1971) suggested exponential decay \( h_{ij} = e^{-\beta_{ij}(t-s)} \).

Compensator

\[
\Lambda_1 : \mu_1 \, T + \frac{\alpha_{11}}{\beta_{11}} \sum_{i=1}^{n} \left(1 - e^{-\beta_1(T-t_i)}\right) + \frac{\alpha_{12}}{\beta_{12}} \sum_{j=1}^{m} \left(1 - e^{-\beta_1(T-t_j)}\right)
\]

and similarly for \( \Lambda_2 \).
First Aspect: Covariance Estimation
Second Aspect: Expected Return Estimation
Third Aspect: Cardinally Constrained Optimal Portfolio

Contribution

Current Practice
Marked Point Process with Limit Orderbook, Shek (2009)

Bivariate Hawkes
Simulated result

Figure: Simulated intensity of a bivariate Hawkes process
Unconditional Arrival Intensities

Intensities at layer zero

Figure: Estimated intensities using overlapping 1min windows.
First Aspect: Covariance Estimation
Second Aspect: Expected Return Estimation
Third Aspect: Cardinally Constrained Optimal Portfolio

Current Practice
Marked Point Process with Limit Orderbook, Shek (2009)

Order Book Evolution

Figure: Snapshots of a 10-layer deep LOB at $\tau$ (gray lines) and $\tau^+$ (black lines) for BP.
Marked Bivariate Hawkes

Probability of order completion

Figure: Probability of order completion within 5 seconds from submission. Dots are empirical data and line is the fitted function $0.935x^{-0.155}$. 
Marked Bivariate Hawkes
Marks that incorporates orderbook volume

\[
\begin{align*}
\lambda_1 t &= \mu_1 \bar{v}_1 t + \sum_{t_i < t} \alpha_{11} e^{-\beta_{11}(t-t_i)} + \sum_{t_j < t} \alpha_{12} e^{-\beta_{12}(t-t_j)} \\
\lambda_2 t &= \mu_2 \bar{v}_2 t + \sum_{t_i < t} \alpha_{21} e^{-\beta_{22}(t-t_i)} + \sum_{t_j < t} \alpha_{22} e^{-\beta_{21}(t-t_j)}
\end{align*}
\]

where the probability weighted volume

\[
\bar{v}(t, \tau, L; i) = \frac{1}{\sum_{i,l} v_{t,l;i}} \sum_{l=0}^{L} v_{t,l;i} p_{l,i,\tau}
\]

\(p_{l,i,\tau} = \mathbb{P}(t_f < t + \tau|l,i)\) is the probability of order type \(i \in \{1, 2\}\) submitted at layer \(l\) getting completely filled within \(t + \tau\) seconds. \(v_{t,l;i}\) is the cumulative size of orders at time \(t\), at the \(l\)-th layer and of type \(i\).
<table>
<thead>
<tr>
<th></th>
<th>Bivariate unmarked</th>
<th>Bivariate marked with $\bar{v}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_1$</td>
<td>0.017</td>
<td>0.041</td>
</tr>
<tr>
<td>$\mu_2$</td>
<td>0.014</td>
<td>0.044</td>
</tr>
<tr>
<td>$\alpha_{11}$</td>
<td>2.017</td>
<td>2.123</td>
</tr>
<tr>
<td>$\alpha_{22}$</td>
<td>3.314</td>
<td>3.088</td>
</tr>
<tr>
<td>$\alpha_{12}$</td>
<td>0.332</td>
<td>0.260</td>
</tr>
<tr>
<td>$\alpha_{21}$</td>
<td>0.453</td>
<td>0.482</td>
</tr>
<tr>
<td>$\beta_{11}$</td>
<td>7.498</td>
<td>7.760</td>
</tr>
<tr>
<td>$\beta_{22}$</td>
<td>11.405</td>
<td>10.473</td>
</tr>
<tr>
<td>$\beta_{12}$</td>
<td>4.275</td>
<td>3.004</td>
</tr>
<tr>
<td>$\beta_{21}$</td>
<td>8.983</td>
<td>9.784</td>
</tr>
<tr>
<td>$l(\theta)$</td>
<td>-5190</td>
<td>-5184</td>
</tr>
</tbody>
</table>

**Table:** MLE result for marked and unmarked models
Goodness-of-Fit
Assuming a standard Poisson process

Figure: QQ-plot for empirical inter-arrival time for buy orders, referenced against a standard exponential distribution.
Goodness-of-Fit
Based on a marked bivariate Hawkes model

Figure: Fitted quantiles based on bivariate Hawkes model marked with order book weighted volume.
Third Aspect: Optimization with Cardinality Constraint

**Cardinally Constrained Quadratic Program (CCQP)**

\[
\begin{align*}
    f^* &= \min_x c^\top x + \frac{1}{2} x^\top H x \\
    \text{s.t.} \quad A x &\geq b \\
    \sum_i 1 \{x_i \neq 0\} &= K
\end{align*}
\]
Current Methodologies

- simplification of the objection function, such as linearization of the quadratic term (e.g. MAD), leading to simpler mixed-integer linear programs.
- heuristic search methods based on genetic algorithm, tabu search or simulated annealing.
- branch-and-cut and branch-and-bound methods that solve relaxations of the CCQP and partition the feasible region to avoid fractional integer solutions.
Projection onto PCA Space
Factor space projection, clustering, local search

Given a return matrix, \( R \in \mathbb{R}^{T \times N} \), we aim to effectively project this universe of \( N \) stock returns onto an orthogonal \( k \)-dimensional space where often \( k \ll N \). We can write

\[
R = P_k V_k^\top + U
\]

where \( P_k \in \mathbb{R}^{T \times k} \) is the return of our \( k \) factors, \( V_k \in \mathbb{R}^{k \times N} \) is the factor loadings and \( U \in \mathbb{R}^{T \times N} \) is the matrix of specific returns.
Projection onto PCA Space

*K-Means* clustering in reduced PCA space

**Figure**: Projection of 3000 US traded assets onto a 4-dimensional PCA space. 15 clusters are identified and each labeled with different symbol and color.
Local Relaxation
Local relaxation within clusters

- Define measure of distance as function of risk tolerance
- Project returns using PCA
- Identify K (the cardinality) clusters with *K-means*
- Follow iterative steps
  - Solve a local relaxation of the QP by including only a subset of cluster members (nearest neighbors to the current solution).
  - Identify new centroid using “center of gravity”
  - Dynamically (i) adjust cluster size based on new centroid weight, (ii) replace centroid if at local minimum
Iterative Relocation of Centroids

New centroids identified by center of gravity

**Figure:** Schematic of local relaxation of centroids.
Benchmark
DLR, Successive Truncation, CPLEX MIQP

- Successive truncation iteratively discard portion of assets that has small weight
- CPLEX: C++ API, parallel 8 Intel Core i7 2.8GHz CPUs, 16GB RAM
- Local Relaxation: R implementation, single CPU
First Aspect: Covariance Estimation
Second Aspect: Expected Return Estimation
Third Aspect: Cardinaly Constrained Optimal Portfolio

Contribution

Current Practice
Local Smoothing Algorithm, Murray & Shek (2010)

Performance
500 Assets: DLR significantly faster than CPLEX MIQP

Figure: Comparison of Local Relaxation vs CPLEX, 500 assets, K = 15.
Performance
3000 Assets: DLR significantly faster than CPLEX MIQP
Cardinally Constrained Portfolio Optimization
Putting everything back together

\[ f^* = \min_x c^\top x + \frac{1}{2} x^\top Hx \]
\[ \text{s.t.} \quad Ax \geq b \]
\[ \sum_i 1_{\{x_i \neq 0\}} = K \]

This work has tackled these key aspects of optimal portfolios:

- modeling of expected return, \( c \); covariance matrix, \( H \);
- proposed optimization framework to solve resulting cardinality constrained portfolio.
RealizedGarch (Hansen, Huang & Shek, 2009; *submitted to Journal of Financial Econometrics*) framework - incorporates high frequency data to give multiperiod forecast of variance and covariance.

Marked point process (Shek, 2010; *submitted to Journal of Financial Markets*) - incorporates order book information to model market order submission intensity.

Local relaxation search algorithm (Murray & Shek, 2010; *working paper*) - uses projection and clustering to solve cardinality constrained portfolio optimization.
First Aspect: Covariance Estimation
Second Aspect: Expected Return Estimation
Third Aspect: Cardinaly Constrained Optimal Portfolio

Contribution

Thank you.
Hansen, P., Huang, Z., Shek, H.
Realized GARCH: A Complete Model of Returns and Realized Measures of Volatility.

Shek, H.

Murray, W., Shek, H.
Dynamic Local Relaxation Method for Cardinality Constrained Portfolio Optimization.
2010. Working Paper