# Vast Volatility Matrix Estimation using High Frequency Data for Portfolio Selection

**Jianqing Fan**

**Princeton** University

With **Yingying Li and Ke Yu**

http://www.princeton.edu/∼jqfan

April 23, 2010

# About this talk

- How to select sparsely optimal portfolio?
- How to use **high-frequency** data to shorten time horizon?
- How large the universe of assets can be handled?
- How does the estimation of **vast** covariance matrix impact on the allocation vector and portfolio risk?

# About this talk

- How to select sparsely optimal portfolio?
- How to use **high-frequency** data to shorten time horizon?
- How large the universe of assets can be handled?
- How does the estimation of **vast** covariance matrix impact on the allocation vector and portfolio risk?

# About this talk

- How to select sparsely optimal portfolio?
- How to use **high-frequency** data to shorten time horizon?
- How large the universe of assets can be handled?
- How does the estimation of **vast** covariance matrix impact on the allocation vector and portfolio risk?

# About this talk

- How to select sparsely optimal portfolio?
- How to use **high-frequency** data to shorten time horizon?
- How large the universe of assets can be handled?
- How does the estimation of **vast** covariance matrix impact on the allocation vector and portfolio risk?

# Outline

1. Introduction
2. Portfolio selection with time-varying covariance.
3. Covariance estimation based high-frequency data
4. An empirical study
5. A simulation study

# Introduction

# Markowitz's Mean-variance analysis

**Portfolio allocation**: $\min_{\mathbf{w}^T\mathbf{1}=1, \mathbf{w}^T\mu=r_0} \mathbf{w}^T\Sigma\mathbf{w}$

Solution: $\mathbf{w} = c_1\,\Sigma^{-1}\mu + c_2\,\Sigma^{-1}\mathbf{1}$

- ★ Cornerstone of modern finance.
- ★ Too **sensitive** on input vectors and their estimation errors.
- ★ More severe for large portfolios: 2000 stocks involves **2 m** parameters! Error accumulation can be huge.
- Impact of dimensionality is large:

  Risk: $\mathbf{w}^T\hat{\Sigma}\mathbf{w}$. Allocation: $\hat{c}_1\hat{\Sigma}^{-1}\mathbf{1} + \hat{c}_2\hat{\Sigma}^{-1}\hat{\mu}$.

# Markowitz's Mean-variance analysis

**Portfolio allocation**: $\min_{\mathbf{w}^T\mathbf{1}=1, \mathbf{w}^T\mu=r_0} \mathbf{w}^T\Sigma\mathbf{w}$

Solution: $\mathbf{w} = c_1\,\Sigma^{-1}\mu + c_2\,\Sigma^{-1}\mathbf{1}$

★ Cornerstone of modern finance.

★ Too **sensitive** on input vectors and their estimation errors.

★ More severe for large portfolios: 2000 stocks involves **2 m** parameters! Error accumulation can be huge.

- Impact of dimensionality is large:

  Risk: $\mathbf{w}^T\hat{\Sigma}\mathbf{w}$. Allocation: $\hat{c}_1\hat{\Sigma}^{-1}\mathbf{1} + \hat{c}_2\hat{\Sigma}^{-1}\hat{\mu}$.

# Markowitz's Mean-variance analysis

**Portfolio allocation**: $\min_{\mathbf{w}^T\mathbf{1}=1, \mathbf{w}^T\mu=r_0} \mathbf{w}^T\Sigma\mathbf{w}$

Solution: $\mathbf{w} = c_1\,\Sigma^{-1}\mu + c_2\,\Sigma^{-1}\mathbf{1}$

★ Cornerstone of modern finance.

★ Too **sensitive** on input vectors and their estimation errors.

★ More severe for large portfolios: 2000 stocks involves **2 m** parameters! Error accumulation can be huge.

- Impact of dimensionality is large:

  Risk: $\mathbf{w}^T\hat{\Sigma}\mathbf{w}$.  Allocation: $\hat{c}_1\hat{\Sigma}^{-1}\mathbf{1} + \hat{c}_2\hat{\Sigma}^{-1}\hat{\mu}$.

# Markowitz's Mean-variance analysis

**Portfolio allocation**: $\min_{\mathbf{w}^T\mathbf{1}=1, \mathbf{w}^T\mu=r_0} \ \mathbf{w}^T\Sigma\mathbf{w}$

Solution: $\mathbf{w} = c_1\,\Sigma^{-1}\mu + c_2\,\Sigma^{-1}\mathbf{1}$

★ Cornerstone of modern finance.

★ Too **sensitive** on input vectors and their estimation errors.

★ More severe for large portfolios: 2000 stocks involves **2 m** parameters! Error accumulation can be huge.

■ Impact of dimensionality is large:

Risk: $\mathbf{w}^T\hat{\Sigma}\mathbf{w}$. Allocation: $\hat{c}_1\hat{\Sigma}^{-1}\mathbf{1} + \hat{c}_2\hat{\Sigma}^{-1}\hat{\mu}$.

# Exposure-constrained portfolio selection

**Portfolio allocation**: (Fan, et al, 08; DeMiguel et al, 08; Bordie et al, 08)

$$\min_{\mathbf{w}^T\mathbf{1}=1,\ \ \mathbf{A}\mathbf{w}=\mathbf{a}} \mathbf{w}^T\Sigma\mathbf{w}, \qquad \|\mathbf{w}\|_1 \le \mathbf{c}.$$

**Constraints**:

- expected return or sector exposures via **A**.
- **short positions**: $\mathbf{w}^- \le (c-1)/2$,

  since $w^+ + w^- \le c, \quad w^+ - w^- = 1.$

  $c = 1 \Longrightarrow$ no short-sale; $c = \infty \Longrightarrow$ Markowitz problem.

**Portfolio selection**: solution is usually sparse.

**Applicability**: Any coherent risk measures (Artzner et al, 1999)

# Exposure-constrained portfolio selection

**Portfolio allocation**: (Fan, et al, 08; DeMiguel et al, 08; Bordie et al, 08)

$$\min_{\mathbf{w}^T\mathbf{1}=1,\ \mathbf{A}\mathbf{w}=\mathbf{a}} \mathbf{w}^T\Sigma\mathbf{w}, \qquad \|\mathbf{w}\|_1 \le \mathbf{c}.$$

**Constraints**:

- expected return or sector exposures via **A**.
- **short positions**: $\mathbf{w}^- \le (c-1)/2$,

  since $w^+ + w^- \le c, \quad w^+ - w^- = 1.$

  $c = 1 \Longrightarrow$ no short-sale; $c = \infty \Longrightarrow$ Markowitz problem.

**Portfolio selection**: solution is usually sparse.

**Applicability**: Any coherent risk measures (Artzner et al, 1999)

# Exposure-constrained portfolio selection

**Portfolio allocation**: (Fan, et al, 08; DeMiguel et al, 08; Bordie et al, 08)

$$\min_{\mathbf{w}^T\mathbf{1}=1,\ \ \mathbf{A}\mathbf{w}=\mathbf{a}} \mathbf{w}^T\Sigma\mathbf{w}, \qquad \|\mathbf{w}\|_1 \leq \mathbf{c}.$$

**Constraints**:

- expected return or sector exposures via **A**.
- **short positions**: $\mathbf{w}^- \leq (c-1)/2$,

  since $w^+ + w^- \leq c, \quad w^+ - w^- = 1.$

  $c = 1 \Longrightarrow$ no short-sale; $c = \infty \Longrightarrow$ Markowitz problem.

**Portfolio selection**: solution is usually sparse.

**Applicability**: Any coherent risk measures (Artzner et al, 1999)

# Utility Approximations

Utility Approx.: Let $M(\mu,\Sigma) = \mathbf{w}^T\mu - \lambda\mathbf{w}^T\Sigma\mathbf{w}$ be expected utility.

$$\begin{aligned} |M(\hat\mu,\hat\Sigma) - M(\mu,\Sigma)| &\le \|\hat\mu-\mu\|_\infty\|\mathbf{w}\|_1 + \lambda|\hat\Sigma-\Sigma|_\infty\|\mathbf{w}\|_1^2 \\ &\le \|\hat\mu-\mu\|_\infty c + \lambda|\hat\Sigma-\Sigma|_\infty c^2, \end{aligned}$$

- No noise accumulation effect for moderate $c \le 3$, say.
- applicable to any number of assets $p$

Risk Approx.: Letting $R(\mathbf{w},\Sigma) = \mathbf{w}^T\Sigma\mathbf{w}$,

$$|R(\mathbf{w},\hat\Sigma) - R(\mathbf{w},\Sigma)| \le |\hat\Sigma-\Sigma|_\infty c^2,$$

# Utility Approximations

**Utility Approx.**: Let $M(\mu,\Sigma) = \mathbf{w}^T\mu - \lambda\mathbf{w}^T\Sigma\mathbf{w}$ be expected utility.

$$\begin{aligned} |M(\hat{\mu},\hat{\Sigma}) - M(\mu,\Sigma)| &\le \|\hat{\mu}-\mu\|_\infty\|\mathbf{w}\|_1 + \lambda|\hat{\Sigma}-\Sigma|_\infty\|\mathbf{w}\|_1^2 \\ &\le \|\hat{\mu}-\mu\|_\infty c + \lambda|\hat{\Sigma}-\Sigma|_\infty c^2, \end{aligned}$$

- No noise accumulation effect for moderate $c \le 3$, say.
- applicable to any number of assets $p$

**Risk Approx.**: Letting $R(\mathbf{w},\Sigma) = \mathbf{w}^T\Sigma\mathbf{w}$,

$$|R(\mathbf{w},\hat{\Sigma}) - R(\mathbf{w},\Sigma)| \le |\hat{\Sigma}-\Sigma|_\infty c^2,$$

# Risk Approximation Theory

**Actual and Empirical risks**: $R(\mathbf{w}) = \mathbf{w}^T \Sigma \mathbf{w}, \qquad R_n(\mathbf{w}) = \mathbf{w}^T \hat{\Sigma} \mathbf{w}.$

■Theoretical and empirical allocation vector:

$$\mathbf{w}_{opt} = \mathrm{argmin}_{||\mathbf{w}||_1 \le c} R(\mathbf{w}), \qquad \hat{\mathbf{w}}_{opt} = \mathrm{argmin}_{||\mathbf{w}||_1 \le c} \mathrm{R}_n(\mathbf{w})$$

■Risks: $\sqrt{R(\mathbf{w}_{opt})}$ —oracle, $\sqrt{R_n(\hat{\mathbf{w}}_{opt})}$ —empirical; $\sqrt{R(\hat{\mathbf{w}}_{opt})}$ —actual risk of a selected portfolio.

**Theorem 1**: Let $a_n = |\hat{\Sigma} - \Sigma|_\infty$. Then, we have

$$\begin{aligned} |R(\hat{\mathbf{w}}_{opt}) - R(\mathbf{w}_{opt})| &\le 2a_n c^2 \\ |R(\hat{\mathbf{w}}_{opt}) - R_n(\hat{\mathbf{w}}_{opt})| &\le a_n c^2 \\ |R(\mathbf{w}_{opt}) - R_n(\hat{\mathbf{w}}_{opt})| &\le a_n c^2. \end{aligned}$$

# Risk Approximation Theory

**Actual and Empirical risks**: $R(\mathbf{w}) = \mathbf{w}^T\Sigma\mathbf{w}, \qquad R_n(\mathbf{w}) = \mathbf{w}^T\hat{\Sigma}\mathbf{w}.$

■ Theoretical and empirical allocation vector:

$$\mathbf{w}_{opt} = \mathrm{argmin}_{||\mathbf{w}||_1 \le c} R(\mathbf{w}), \qquad \hat{\mathbf{w}}_{opt} = \mathrm{argmin}_{||\mathbf{w}||_1 \le c} \mathrm{R}_n(\mathbf{w})$$

■ Risks: $\sqrt{R(\mathbf{w}_{opt})}$ —oracle, $\sqrt{R_n(\hat{\mathbf{w}}_{opt})}$ —empirical; $\sqrt{R(\hat{\mathbf{w}}_{opt})}$ —actual risk of a selected portfolio.

**Theorem 1**: Let $a_n = |\hat{\Sigma} - \Sigma|_\infty$. Then, we have

$$\begin{aligned} |R(\hat{\mathbf{w}}_{opt}) - R(\mathbf{w}_{opt})| &\le 2a_n c^2 \\ |R(\hat{\mathbf{w}}_{opt}) - R_n(\hat{\mathbf{w}}_{opt})| &\le a_n c^2 \\ |R(\mathbf{w}_{opt}) - R_n(\hat{\mathbf{w}}_{opt})| &\le a_n c^2. \end{aligned}$$

# Risk Approximation Theory

**Actual and Empirical risks**: $R(\mathbf{w}) = \mathbf{w}^T \Sigma \mathbf{w}, \qquad R_n(\mathbf{w}) = \mathbf{w}^T \hat{\Sigma} \mathbf{w}.$

■Theoretical and empirical allocation vector:

$$\mathbf{w}_{opt} = \operatorname{argmin}_{||\mathbf{w}||_1 \le c} R(\mathbf{w}), \qquad \hat{\mathbf{w}}_{opt} = \operatorname{argmin}_{||\mathbf{w}||_1 \le c} \mathrm{R}_n(\mathbf{w})$$

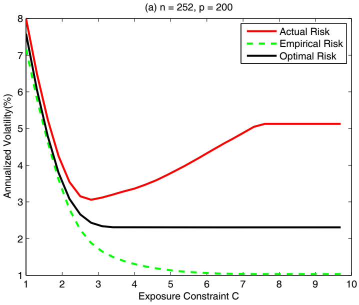
■Risks: $\sqrt{R(\mathbf{w}_{opt})}$ —oracle, $\sqrt{R_n(\hat{\mathbf{w}}_{opt})}$ —empirical; $\sqrt{R(\hat{\mathbf{w}}_{opt})}$ —actual risk of a selected portfolio.

**Theorem 1**: Let $a_n = |\hat{\Sigma} - \Sigma|_\infty$. Then, we have

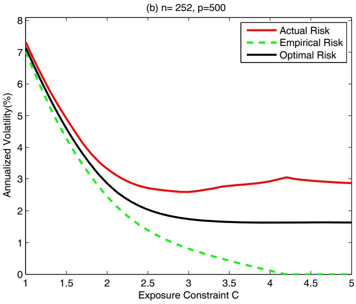
$$\begin{aligned} |R(\hat{\mathbf{w}}_{opt}) - R(\mathbf{w}_{opt})| &\le 2a_n c^2 \\ |R(\hat{\mathbf{w}}_{opt}) - R_n(\hat{\mathbf{w}}_{opt})| &\le a_n c^2 \\ |R(\mathbf{w}_{opt}) - R_n(\hat{\mathbf{w}}_{opt})| &\le a_n c^2. \end{aligned}$$

# Impact of dimensionality

## Actual vs Empirical risks




## Actual vs Empirical




**Theorem 2**: If $\max_{i,j} P\{\sqrt{n}|\sigma_{ij} - \hat{\sigma}_{ij}| > x\} < \exp(-\mathbf{C}\mathbf{x}^{1/a})$ for large $x$,

$$|\Sigma - \hat{\Sigma}|_\infty = O_P\left(\frac{(\log \mathbf{p})^{\mathbf{a}}}{\sqrt{\mathbf{n}}}\right).$$

- Impact of dimensionality is limited.
- on inverse of tail.

# Portfolio Selection with dynamic covariance

# Time-dependent volatility matrix

**Return and Risk** with holding period $\tau$:

$$\textbf{Return} = \mathbf{w}^T \mathbf{R}_{t,\tau} = \mathbf{w}^T \int_t^{t+\tau} d\mathbf{X}_s, \qquad \textbf{risk} = \mathbf{w}^T \Sigma_{t,\tau} \mathbf{w},$$

where $\Sigma_{t,\tau} = E_t \int_t^{t+\tau} \mathbf{S}_u du$, allowing **stochastic** volatility and $\mathbf{S}_u = \left( \sigma_u^{(i)} \sigma_u^{(j)} \rho_u^{(i,j)} \right)$ is instantaneous cov matrix.

**Portfolio allocation and selection**:

$$\min_{\mathbf{w}^T \mathbf{1} = 1, \ \mathbf{A}\mathbf{w} = \mathbf{a}} \mathbf{w}^T \Sigma_{t,\tau} \mathbf{w}, \qquad \|\mathbf{w}\|_1 \le \mathbf{c}.$$

# Time-dependent volatility matrix

**Return and Risk** with holding period $\tau$:

$$\textbf{Return} = \mathbf{w}^T \mathbf{R}_{t,\tau} = \mathbf{w}^T \int_t^{t+\tau} d\mathbf{X}_s, \qquad \textbf{risk} = \mathbf{w}^T \Sigma_{t,\tau} \mathbf{w},$$

where $\Sigma_{t,\tau} = E_t \int_t^{t+\tau} \mathbf{S}_u du$, allowing **stochastic** volatility and $\mathbf{S}_u = \left(\sigma_u^{(i)} \sigma_u^{(j)} \rho_u^{(i,j)}\right)$ is instantaneous cov matrix.

**Portfolio allocation and selection**:

$$\min_{\mathbf{w}^T \mathbf{1}=1, \ \ \mathbf{A}\mathbf{w}=\mathbf{a}} \mathbf{w}^T \Sigma_{t,\tau} \mathbf{w}, \qquad \|\mathbf{w}\|_1 \le \mathbf{c}.$$

# Prediction of Covariance Matrix

Covariance matrix is predicted based on following approximations:

**short-horizon $\tau$** : $\frac{1}{\tau}\Sigma_{t,\tau} \approx \frac{1}{h}\int_{t-h}^{t} \mathbf{S}_u du$ (use of continuity)

**long-horizon $\tau$** : $\frac{1}{\tau}\Sigma_{t,\tau} \approx \frac{1}{h}E\int_{t-h}^{t} \mathbf{S}_u du$ (use of ergoticity)

- Even with observed $\mathbf{S}_u$ in the past, $\Sigma_{t,\tau}$ is at best approximated.
- Important to reduce the sensitivity of $\mathbf{w}$ on the prediction of $\Sigma_{t,\tau}$.
- Gross-exposure constraint is an effective method.

# Prediction of Covariance Matrix

Covariance matrix is predicted based on following approximations:

**short-horizon $\tau$** : $\frac{1}{\tau}\Sigma_{t,\tau} \approx \frac{1}{h}\int_{t-h}^{t} \mathbf{S}_u du$ (use of continuity)

**long-horizon $\tau$** : $\frac{1}{\tau}\Sigma_{t,\tau} \approx \frac{1}{h}E\int_{t-h}^{t} \mathbf{S}_u du$ (use of ergoticity)

- Even with observed $\mathbf{S}_u$ in the past, $\Sigma_{t,\tau}$ is at best approximated.
- Important to reduce the sensitivity of $\mathbf{w}$ on the prediction of $\Sigma_{t,\tau}$.
- Gross-exposure constraint is an effective method.

# High- and low-frequency data

**Low frequency Data**: Daily data w/ $h = 252$ or $h = 512$ days.

- Estimated is the expected covariance matrix from $[t-h, t]$.
- Can be very different from $\Sigma_{t,\tau}$ next day or week.
- Not applicable to **short** holding period.
- Applicable to long holding period only when stationary.

**Use of high-frequency data**:

- ★ More data available for estimating covariance matrix
- ★ Shorten the time interval, reducing approximation errors
- ★ Adapts better local correlation.
- ★ Applicable to both **long- and short-term** holding periods

# High- and low-frequency data

**Low frequency Data**: Daily data w/ $h = 252$ or $h = 512$ days.

- Estimated is the expected covariance matrix from $[t-h, t]$.
- Can be very different from $\Sigma_{t,\tau}$ next day or week.
- Not applicable to **short** holding period.
- Applicable to long holding period only when stationary.

**Use of high-frequency data**:

★ More data available for estimating covariance matrix

★ Shorten the time interval, reducing approximation errors

★ Adapts better local correlation.

★ Applicable to both **long- and short-term** holding periods

# Covariance Estimation

# Using High-Frequency Data

# Style features

- Microstructure noise (Aït-Sahalia, Mykland, Zhang, RFS, 05);
- Nonsynchronized trading (Barndorff-Nielsen, Hansen, Lunde and Shephard, EconJ,08);
- Jumps in the data (Fan and Wang, 07; BNS, 04, 06, JFEC);
- Data cleaning (BNHLS, EconJ, 09)

# Integrated volatility: Diagonal elements

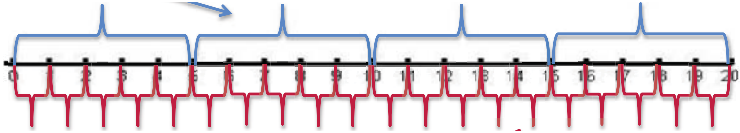
**Model**: $Y_{t_i} = X_{t_i} + \varepsilon_{t_i}$, $X_{t_i}$ —latent log-price, $\eta^2 = \mathrm{var}(\varepsilon)$

- Two-scale and Multi-scaled realized volatility. (AMZ, 05; Zhang, 07)
- Realized kernel method (BNHLS, JFEC 09, JEcon, 09)
- Wavelets (Fan and Wang, 07) and Bipower (BNS, 04, 06, JFEC)
- Quasi-MLE (Xiu, 09)
- Pre-averaging (smoothing) (Jacod, Li, Mykland, Podolskij, Vetter, 09).

# Sub-sampling

**Subsampling**: Use once every $K$ points

$$
\begin{aligned}
RV_{K,i} &= \sum_{j=1}^{n_s} (Y_{t_{i+jK}} - Y_{t_{i+(j-1)K}})^2, \qquad n_s = n/K, \quad \Theta = \int_{t-h}^{t} \sigma_u^2 du. \\
&= \Theta + 2n_s\eta^2 + \left[4n_s E\varepsilon^4 + \frac{2}{n_s}\int \sigma_t^4 dt\right]^{1/2} \cdot N(0,1),
\end{aligned}
$$




**Averaging** : 
$$
\begin{aligned}
[Y]^{(K)} &= \frac{1}{K}\sum_{i=0}^{K-1} R_{K,i} = \frac{1}{K}\sum_{i=1}^{n-K} (Y_{t_i+K} - Y_{t_i})^2 \\
&\approx \Theta + 2n_s\eta^2 + \left[\frac{4n_s}{K} E\varepsilon^4 + \frac{4}{3n_s}\int \sigma_t^4 dt\right]^{1/2} \cdot N(0,1)
\end{aligned}
$$

# Two-scale Realized Volatility

**TSRV**: $[Y]^{(K)} - [Y]^{(1)}/K \cdot \frac{n-K+1}{n}$

**Asymptotic normality** (AMZ, 05): with optimal choice $K = cn^{2/3}$,

$$n^{1/6}(TSRV - \Theta) \to \left[8c^{-2}\eta^4 + c\frac{4}{3}\int \sigma_t^4 dt\right]^{1/2} \cdot N(0,1).$$

**Theorem 3** (Concentration inequality): For large $x$ that satisfies $|x| \leq cn^{1/6}$,

$$P\{n^{1/6}|TSRV - \Theta| > x\} \leq 3\exp\{-Cx^2\}$$

- By Thm 2, diagonals be estimated uniformly with rate $O(\frac{(\log p)^{1/2}}{n_{\min}^{1/6}})$.

# Two-scale Realized Volatility

**TSRV**: $[Y]^{(K)} - [Y]^{(1)}/K \cdot \frac{n-K+1}{n}$

**Asymptotic normality** (AMZ, 05): with optimal choice $K = cn^{2/3}$,

$$n^{1/6}(TSRV - \Theta) \rightarrow \left[8c^{-2}\eta^4 + c\frac{4}{3}\int \sigma_t^4 dt\right]^{1/2} \cdot N(0,1).$$

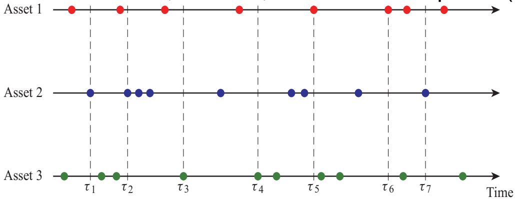
**Theorem 3** (Concentration inequality): For large $x$ that satisfies $|x| \leq cn^{1/6}$,

$$P\{n^{1/6}|TSRV - \Theta| > x\} \leq 3\exp\{-Cx^2\}$$

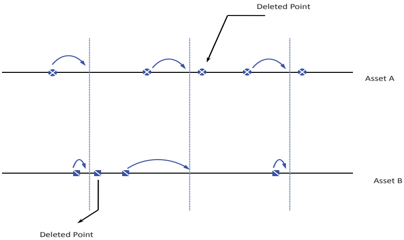
■ By Thm 2, diagonals be estimated uniformly with rate $O(\frac{(\log p)^{1/2}}{n_{\min}^{1/6}})$.

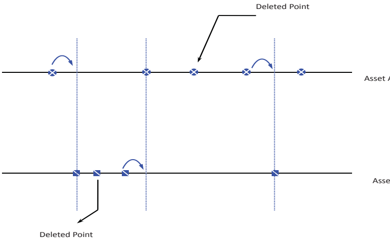
# Data Synchronization

**Refresh time**: Barndorff-Nielsen, Hansen, Lunde and Shephard (2008)




**Previous ticks and its generalization**: $\{\tau_i - \tau_{i-1}\}$ are i.i.d. $O_P(n^{-1})$, and at least 1 data for **each asset** in $(\tau_{i-1}, \tau_i]$.



# Estimation of integrated covariance

1. Two-Scale Realized Covariance (Zhang, 09):

$$\text{TSCV} = [Y_1, Y_2]^{(K)} - [Y_1, Y_2]^{(1)}/K \cdot \frac{\tilde{n} - K + 1}{\tilde{n}},$$

   where $\tilde{n}$ is no of synchronized data, and

   $[Y_1, Y_2]^{(K)} = \frac{1}{K}\sum_{i=K}^{\tilde{n}}(Y_{1,t_i} - Y_{1,t_{i-K}})(Y_{2,t_i} - Y_{2,t_{i-K}})$, **subsam cov**

2. Realized Covariance(BNHLS, 08): log-return $\mathbf{y}_t$

$$K(X) = \sum_{h=-H}^{H} k\left(\frac{h}{H+1}\right)\Gamma_h, \qquad \Gamma(h) = \sum_{j=|h|+1}^{n} \mathbf{y}_j \mathbf{y}'_{j-|h|}$$

3. QMLE (Aït-Sahalia, Fan and Xiu, 2010)

$$\widehat{\langle Y_1, Y_2 \rangle} = \frac{1}{4}\{\widehat{\langle Y_1 + Y_2, Y_1 + Y_2 \rangle}_{QMLE} - \widehat{\langle Y_1 - Y_2, Y_1 - Y_2 \rangle}_{QMLE}\}$$

# Estimation of integrated covariance

1. Two-Scale Realized Covariance (Zhang, 09):

$$\text{TSCV} = [Y_1, Y_2]^{(K)} - [Y_1, Y_2]^{(1)}/K \cdot \frac{\tilde{n} - K + 1}{\tilde{n}},$$

where $\tilde{n}$ is no of synchronized data, and

$[Y_1, Y_2]^{(K)} = \frac{1}{K}\sum_{i=K}^{\tilde{n}}(Y_{1,t_i} - Y_{1,t_{i-K}})(Y_{2,t_i} - Y_{2,t_{i-K}})$, **subsam cov**

2. Realized Covariance(BNHLS, 08): log-return $\mathbf{y}_t$

$$K(X) = \sum_{h=-H}^{H} k\left(\frac{h}{H+1}\right)\Gamma_h, \qquad \Gamma(h) = \sum_{j=|h|+1}^{n} \mathbf{y}_j \mathbf{y}_{j-|h|}'$$

3. QMLE (Aït-Sahalia, Fan and Xiu, 2010)

$$\widehat{\langle Y_1, Y_2 \rangle} = \frac{1}{4}\{\widehat{\langle Y_1 + Y_2, Y_1 + Y_2\rangle}_{QMLE} - \widehat{\langle Y_1 - Y_2, Y_1 - Y_2\rangle}_{QMLE}\}$$

# Estimation of integrated covariance

1. Two-Scale Realized Covariance (Zhang, 09):

$$\text{TSCV} = [Y_1, Y_2]^{(K)} - [Y_1, Y_2]^{(1)}/K \cdot \frac{\tilde{n} - K + 1}{\tilde{n}},$$

where $\tilde{n}$ is no of synchronized data, and

$[Y_1, Y_2]^{(K)} = \frac{1}{K} \sum_{i=K}^{\tilde{n}} (Y_{1,t_i} - Y_{1,t_{i-K}})(Y_{2,t_i} - Y_{2,t_{i-K}})$, **subsam cov**

2. Realized Covariance(BNHLS, 08): log-return $\mathbf{y}_t$

$$K(X) = \sum_{h=-H}^{H} k\left(\frac{h}{H+1}\right)\Gamma_h, \qquad \Gamma(h) = \sum_{j=|h|+1}^{n} \mathbf{y}_j \mathbf{y}'_{j-|h|}$$

3. QMLE (Aït-Sahalia, Fan and Xiu, 2010)

$$\widehat{\langle Y_1, Y_2 \rangle} = \frac{1}{4}\{\widehat{\langle Y_1 + Y_2, Y_1 + Y_2 \rangle}_{QMLE} - \widehat{\langle Y_1 - Y_2, Y_1 - Y_2 \rangle}_{QMLE}\}$$

# A concentration inequality for TSCV

**Theorem 4**. For large $x$ that satisfies $|x| \leq c\tilde{n}^{1/6}$,

$$P\{\tilde{n}^{1/6}|\text{TSCV} - \int_0^1 \sigma_t^{Y_1}\sigma_t^{Y_2}\rho_t^{(Y_1,Y_2)}dt| > x\} \leq 3\exp\{-Cx^2\}.$$

**Conditions**:

1. **Log-price**: $dX_t^{(i)} = \sigma_t^{(i)} dB_t^{(i)}$ with $cor(B_t^{(i)}, B_t^{(j)}) = \rho_t^{(i,j)}$.
2. **Volatility**: $|\sigma_t^{(i)}| < C_\sigma$.
3. **Refresh time**: $\sup_j |\tau_j - \tau_{j-1}| \leq C_\Delta / n_1$
4. **Noise**: $\{\varepsilon_{t_j}^{Y_i}\}$ are independent, also independent of $X^{(i)}$.

# A concentration inequality for TSCV

**Theorem 4**. For large $x$ that satisfies $|x| \leq c\tilde{n}^{1/6}$,

$$P\{\tilde{n}^{1/6}|\mathrm{TSCV} - \int_0^1 \sigma_t^{Y_1}\sigma_t^{Y_2}\rho_t^{(Y_1,Y_2)}dt| > x\} \leq 3\exp\{-Cx^2\}.$$

**Conditions**:

1. **Log-price**: $dX_t^{(i)} = \sigma_t^{(i)} dB_t^{(i)}$ with $cor(B_t^{(i)}, B_t^{(j)}) = \rho_t^{(i,j)}$.
2. **Volatility**: $|\sigma_t^{(i)}| < C_\sigma$.
3. **Refresh time**: $\sup_j |\tau_j - \tau_{j-1}| \leq C_\Delta / n_1$
4. **Noise**: $\{\varepsilon_{t_j}^{Y_i}\}$ are independent, also independent of $X^{(i)}$.

# Applications to Portfolio Allocation

# Portfolio Optimization

**Portfolio allocation**: $\min_{\mathbf{w}^T\mathbf{1}=1,\ \ \|\mathbf{w}\|_1\le c} \mathbf{w}^T\widehat{\Sigma}\mathbf{w}$. The actual risk is no larger than $2|\hat{\Sigma}-\Sigma|_\infty c^2$ away from the oracle.

**Estimation of Covariance**

1. **Pairwise refresh**: Componentwise estimation, far more data, but $\hat{\Sigma}$ is **not** semi-positive:

$$|\hat{\Sigma}-\Sigma|_\infty = O\left(\frac{\sqrt{\log p}}{\bar{n}^{1/6}}\right), \qquad \bar{n} = \min_{i,j} n_{i,j}.$$

2. **All refresh**: Far less data, but $\hat{\Sigma}$ is semi-positive:

$$|\hat{\Sigma}-\Sigma|_\infty = O\left(\frac{\sqrt{\log p}}{n_*^{1/6}}\right).$$

# Portfolio Optimization

**Portfolio allocation**: $\min_{\mathbf{w}^T\mathbf{1}=1,\ \|\mathbf{w}\|_1\le c} \mathbf{w}^T\widehat{\Sigma}\mathbf{w}$. The actual risk is no larger than $2|\hat{\Sigma}-\Sigma|_\infty c^2$ away from the oracle.

**Estimation of Covariance**

1. **Pairwise refresh**: Componentwise estimation, far more data, but $\hat{\Sigma}$ is **not** semi-positive:

$$|\hat{\Sigma}-\Sigma|_\infty = O\left(\frac{\sqrt{\log p}}{\bar{n}^{1/6}}\right), \qquad \bar{n} = \min_{i,j} n_{i,j}.$$

2. **All refresh**: Far less data, but $\hat{\Sigma}$ is semi-positive:

$$|\hat{\Sigma}-\Sigma|_\infty = O\left(\frac{\sqrt{\log p}}{n_*^{1/6}}\right).$$

# Portfolio Optimization

**Portfolio allocation**: $\min_{\mathbf{w}^T\mathbf{1}=1,\ \ \|\mathbf{w}\|_1\le c} \mathbf{w}^T\widehat{\Sigma}\mathbf{w}$. The actual risk is no larger than $2|\hat{\Sigma}-\Sigma|_\infty c^2$ away from the oracle.

**Estimation of Covariance**

1. **Pairwise refresh**: Componentwise estimation, far more data, but $\hat{\Sigma}$ is **not** semi-positive:

$$|\hat{\Sigma}-\Sigma|_\infty = O\left(\frac{\sqrt{\log p}}{\bar{n}^{1/6}}\right), \qquad \bar{n} = \min_{i,j} n_{i,j}.$$

2. **All refresh**: Far less data, but $\hat{\Sigma}$ is semi-positive:

$$|\hat{\Sigma}-\Sigma|_\infty = O\left(\frac{\sqrt{\log p}}{n_*^{1/6}}\right).$$

# Projection of symmetric matrices

**Need of projection**: Programming algorithms require $\widehat{\Sigma} \geq 0$.

**Projection 1**: $\mathbf{A}_1^+ = \Gamma^T \text{diag}(\lambda_1^+, \cdots, \lambda_n^+)\Gamma$, for a symmetric matrix with SVD $\mathbf{A} = \Gamma^T \text{diag}(\lambda_1, \cdots, \lambda_n)\Gamma$.

**Projection 2**: $\mathbf{A}_2^+ = (\mathbf{A} - \lambda_{\min}^- I_p)/(1 - \lambda_{\min}^-)$, where $\lambda_{\min}^-$ is the negative part of the minimum eigenvalue.

- Both projections do not alter eigenvectors;
- Applied to the correlation rather than volatility matrix
- The projection has an adverse effect on the performance.

# Projection of symmetric matrices

**Need of projection**: Programming algorithms require $\widehat{\Sigma} \geq 0$.

**Projection 1**: $\mathbf{A}_1^+ = \Gamma^T \mathrm{diag}(\lambda_1^+, \cdots, \lambda_n^+)\Gamma$, for a symmetric matrix with SVD $\mathbf{A} = \Gamma^T \mathrm{diag}(\lambda_1, \cdots, \lambda_n)\Gamma$.

**Projection 2**: $\mathbf{A}_2^+ = (\mathbf{A} - \lambda_{\min}^- I_p)/(1 - \lambda_{\min}^-)$, where $\lambda_{\min}^-$ is the negative part of the minimum eigenvalue.

- Both projections do not alter eigenvectors;
- Applied to the correlation rather than volatility matrix
- The projection has an adverse effect on the performance.

# Projection of symmetric matrices

**Need of projection**: Programming algorithms require $\widehat{\Sigma} \geq 0$.

**Projection 1**: $\mathbf{A}_1^+ = \Gamma^T \mathrm{diag}(\lambda_1^+, \cdots, \lambda_n^+)\Gamma$, for a symmetric matrix with SVD $\mathbf{A} = \Gamma^T \mathrm{diag}(\lambda_1, \cdots, \lambda_n)\Gamma$.

**Projection 2**: $\mathbf{A}_2^+ = (\mathbf{A} - \lambda_{\min}^- I_p)/(1 - \lambda_{\min}^-)$, where $\lambda_{\min}^-$ is the negative part of the minimum eigenvalue.

- Both projections do not alter eigenvectors;
- Applied to the correlation rather than volatility matrix
- The projection has an adverse effect on the performance.

# Remarks

1. It appears projections **distort** more "pairwise refresh" method than "all refresh". Thus, the smaller componentwise estimation errors might not be materialized in implementation.

2. Risk approximation is an upper bound, not necessarily tight.

3. We experimented $2 \times 2$ simulation studies with the first element of $\widehat{\Sigma}$ replaced by its true value. The performance is not always better (about 65%).

4. Because of distortion, pairwise refresh performs not necessarily better.

# Remarks

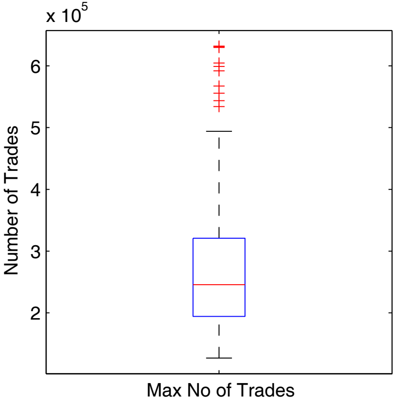
1. It appears projections **distort** more “pairwise refresh” method than “all refresh”. Thus, the smaller componentwise estimation errors might not be materialized in implementation.
2. Risk approximation is an upper bound, not necessarily tight.
3. We experimented $2 \times 2$ simulation studies with the first element of $\widehat{\Sigma}$ replaced by its true value. The performance is not always better (about 65%).
4. Because of distortion, pairwise refresh performs not necessarily better.

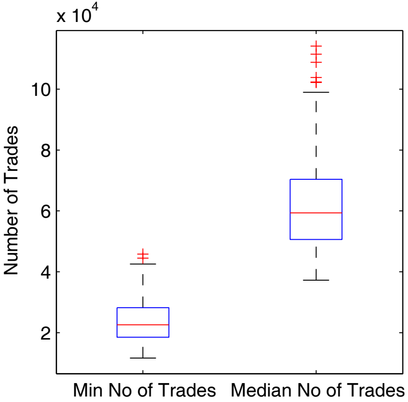
# An Empirical Study

# An empirical testing

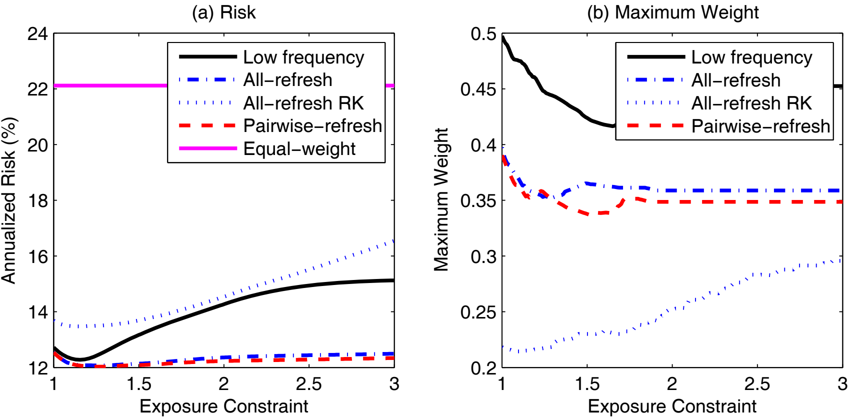
- 30 stocks from DJ Industrial components from 1/2/08–9/30/08 (Total trade: 2,307,004. Average trading: 76,900. Size: 13G)
- Holding period: $\tau = 1$ or 5 days and rebalanced
- testing period: 5/27/08 – 9/30/08 (90 days)
- Risk profile: Use **15 minutes returns** (total $26 * 90 = 2340$ returns), excluding overnight holding risks.
- High frequency $h = 10$ days; low frequency $h = 100$ days

# Summary of Trading Frequencies

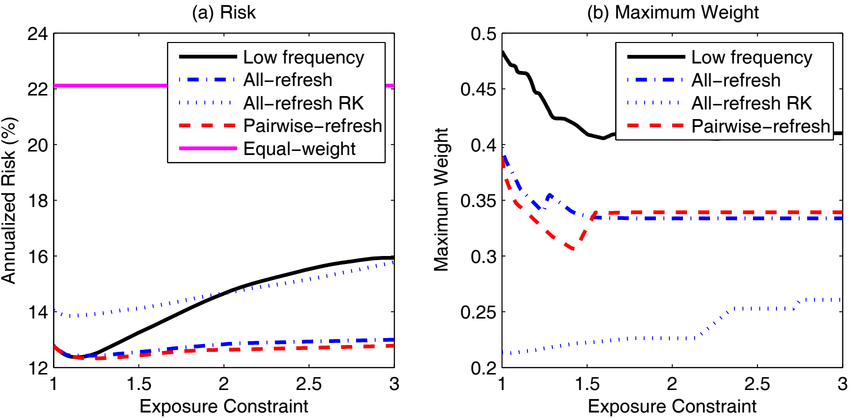


# An empirical result ($\tau = 1$)



# An empirical result ($\tau = 5$)



# A Simulation Study

# Stochastic models

**Log-prices** of $p$-stocks follow the **one-factor** model ($X_0^{(i)} = 1$):

$$dX_t^{(i)} = \mu^{(i)} dt + \rho^{(i)} \sigma_t^{(i)} dB_t^{(i)} + \sqrt{1 - (\rho^{(i)})^2} \sigma_t^{(i)} dW_t + \lambda^{(i)} dZ_t^{(i)},$$

the synchronized data highest freq (second) **—latent** (oracle) price.

**Stochastic volatility**: $\eta_t^{(i)} = \log \sigma_t^{(i)}$ follows Vasicek model (OU):

$$d\eta_t^{(i)} = \alpha^{(i)}(\beta_0^{(i)} - \eta_t^{(i)})dt + \beta_1^{(1)} dB_t^{(i)}.$$

**Choice of parameter**: $\rho^{(i)} = -0.7$, $\lambda^{(i)} = \exp(\beta_0^{(i)})$,

$$(\mu^{(i)}, \beta_0^{(i)}, \beta_1^{(i)}, \alpha^{(i)}) = (0.03, -1, .75, 1/40) \otimes \mathbf{U}^{(i)},$$

where $\mathbf{U}^{(i)} \sim_{i.i.d.} \mathrm{Unif}(0.7, 1.3)^{\otimes 4}$.

# Stochastic models

**Log-prices** of $p$-stocks follow the **one-factor** model $(X_0^{(i)} = 1)$:

$$dX_t^{(i)} = \mu^{(i)} dt + \rho^{(i)} \sigma_t^{(i)} dB_t^{(i)} + \sqrt{1 - (\rho^{(i)})^2} \sigma_t^{(i)} dW_t + \lambda^{(i)} dZ_t^{(i)},$$

the synchronized data highest freq (second) **—latent** (oracle) price.

**Stochastic volatility**: $\eta_t^{(i)} = \log \sigma_t^{(i)}$ follows Vasicek model (OU):

$$d\eta_t^{(i)} = \alpha^{(i)} (\beta_0^{(i)} - \eta_t^{(i)}) dt + \beta_1^{(1)} dB_t^{(i)}.$$

**Choice of parameter**: $\rho^{(i)} = -0.7$, $\lambda^{(i)} = \exp(\beta_0^{(i)})$,

$$(\mu^{(i)}, \beta_0^{(i)}, \beta_1^{(i)}, \alpha^{(i)}) = (0.03, -1, .75, 1/40) \otimes \mathbf{U}^{(i)},$$

where $\mathbf{U}^{(i)} \sim_{i.i.d.} \text{Unif}(0.7, 1.3)^{\otimes 4}$.

# Stochastic models

**Log-prices** of $p$-stocks follow the **one-factor** model ($X_0^{(i)} = 1$):

$$dX_t^{(i)} = \mu^{(i)} dt + \rho^{(i)} \sigma_t^{(i)} dB_t^{(i)} + \sqrt{1 - (\rho^{(i)})^2} \sigma_t^{(i)} dW_t + \lambda^{(i)} dZ_t^{(i)},$$

the synchronized data highest freq (second) **—latent** (oracle) price.

**Stochastic volatility**: $\eta_t^{(i)} = \log \sigma_t^{(i)}$ follows Vasicek model (OU):

$$d\eta_t^{(i)} = \alpha^{(i)} (\beta_0^{(i)} - \eta_t^{(i)}) dt + \beta_1^{(1)} dB_t^{(i)}.$$

**Choice of parameter**: $\rho^{(i)} = -0.7$, $\lambda^{(i)} = \exp(\beta_0^{(i)})$,

$$(\mu^{(i)}, \beta_0^{(i)}, \beta_1^{(i)}, \alpha^{(i)}) = (0.03, -1, .75, 1/40) \otimes \mathbf{U}^{(i)},$$

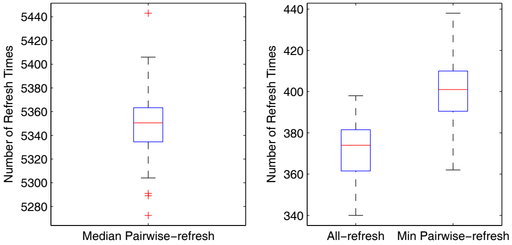
where $\mathbf{U}^{(i)} \sim_{i.i.d.} \text{Unif}(0.7, 1.3)^{\otimes 4}$.

# Stochastic model (II)

**Trading frequency**: Poisson process with $\lambda_i = 0.02i \times 23400$

—no. of seconds / day.

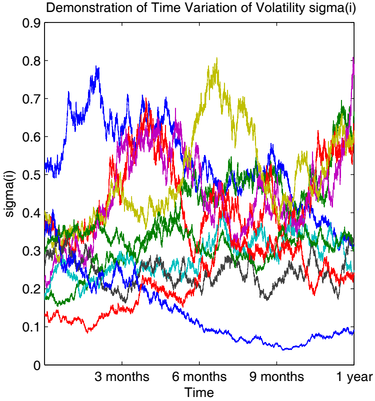
**Size of investment universe**: $p = 50$.

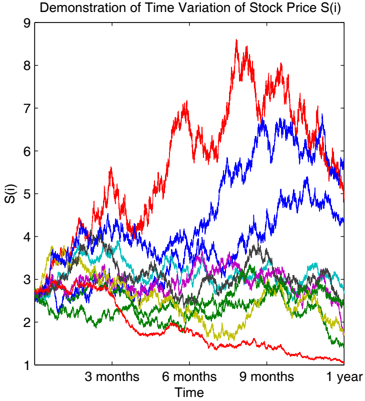



| all-fresh | pairwise-refresh | Ave of min pairwise-refresh |
|---|---|---|
| 375 | 5355 | 410 |

**Microstructural noise**: $Y_{t_{ij}}^{(i)} = X_{t_{ij}}^{(i)} + \mathbf{N(0, 0.0005^2)}$.

# Examples of realized volatilities and prices




- Varying volatility, but relatively calm.

# Risk approximation: In-sample evaluation

**Specific portfolios**: $w_1$ —equal weight, $w_2 = (1, 0, \cdots, 0)^T$,

$$w_3 = (1 + 2/p, -1, 1/p, \cdots, 1/p)^T, \qquad w_3 = (2, -1, 0, \cdots, 0)^T$$

**Evaluation**: Regard risk estimated by Latent price as the true risk.

Median and Robust Standard Deviation (RSD) of Risk

| | Latent | All-refresh TSRV | All-refresh RK | Pairwise TSRV |
|---|---|---|---|---|
| Portfolio | Median(RSD) | Median(RSD) | Median(RSD) | Median(RSD) |
| $w_1$ | 0.440(0.0032) | 0.387 (0.107) | 0.434 (0.024) | 0.419 (0.069) |
| $w_2$ | 0.591(0.0060) | 0.522 (0.125) | 0.623 (0.025) | 0.593 (0.128) |
| $w_3$ | 0.539(0.0044) | 0.469 (0.090) | 0.583 (0.025) | 0.520 (0.073) |
| $w_4$ | 0.844(0.0077) | 0.753 (0.174) | 0.922 (0.041) | 0.839 (0.178) |

# Risk approximation: In-sample evaluation

**Specific portfolios**: $w_1$ —equal weight, $w_2 = (1, 0, \cdots, 0)^T$,

$$w_3 = (1 + 2/p, -1, 1/p, \cdots, 1/p)^T, \qquad w_3 = (2, -1, 0, \cdots, 0)^T$$

**Evaluation**: Regard risk estimated by Latent price as the true risk.

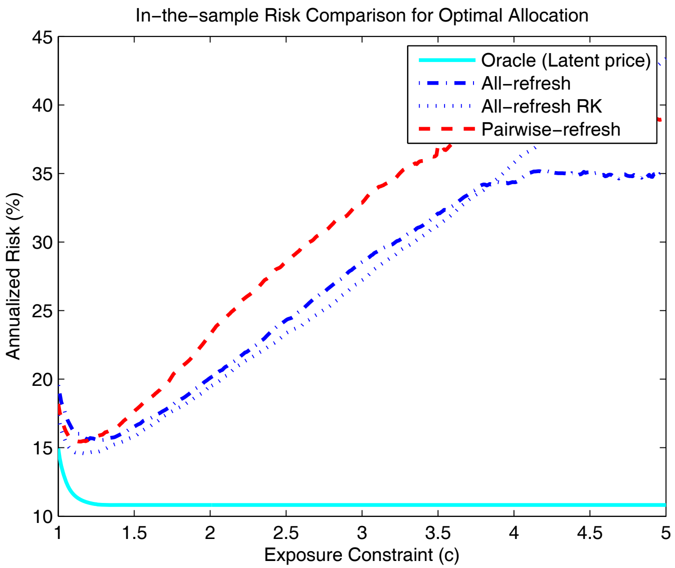
**Median and Robust Standard Deviation (RSD) of Risk**

| | Latent | All-refresh TSRV | All-refresh RK | Pairwise TSRV |
|---|---|---|---|---|
| Portfolio | Median(RSD) | Median(RSD) | Median(RSD) | Median(RSD) |
| $w_1$ | 0.440(0.0032) | 0.387 (0.107) | 0.434 (0.024) | 0.419 (0.069) |
| $w_2$ | 0.591(0.0060) | 0.522 (0.125) | 0.623 (0.025) | 0.593 (0.128) |
| $w_3$ | 0.539(0.0044) | 0.469 (0.090) | 0.583 (0.025) | 0.520 (0.073) |
| $w_4$ | 0.844(0.0077) | 0.753 (0.174) | 0.922 (0.041) | 0.839 (0.178) |

# Risk approximation error

| Median and RSD of Absolute Risk Difference from the Oracle (Latent) | | | |
|---|---|---|---|
| | All-refresh TSRV | All-refresh RK | Pairwise TSRV |
| Portfolio | Median(RSD) | Median(RSD) | Median(RSD) |
| $w_1$ | 0.0889 (0.0769) | 0.0183 (0.0153) | 0.0547 (0.0439) |
| $w_2$ | 0.1054 (0.0700) | 0.0344 (0.0272) | 0.0804 (0.0813) |
| $w_3$ | 0.0936 (0.0665) | 0.0437 (0.0300) | 0.0599 (0.0593) |
| $w_4$ | 0.1470 (0.1022) | 0.0794 (0.0393) | 0.1089 (0.0941) |
| **Median and RSD of $L_1$ Norm of Absolute Covariance Difference ($a_p$)** | | | |
| | All-refresh TSRV | All-refresh RK | Pairwise TSRV |
| Portfolio | Median(RSD) | Median(RSD) | Median(RSD) |
| | 0.2476 (0.1460) | 0.0603 (0.0270) | 0.1730 (0.0746) |

# Evaluation of portfolio allocation: In-sample risk ($\tau = 1$)



# Out-sample evaluation

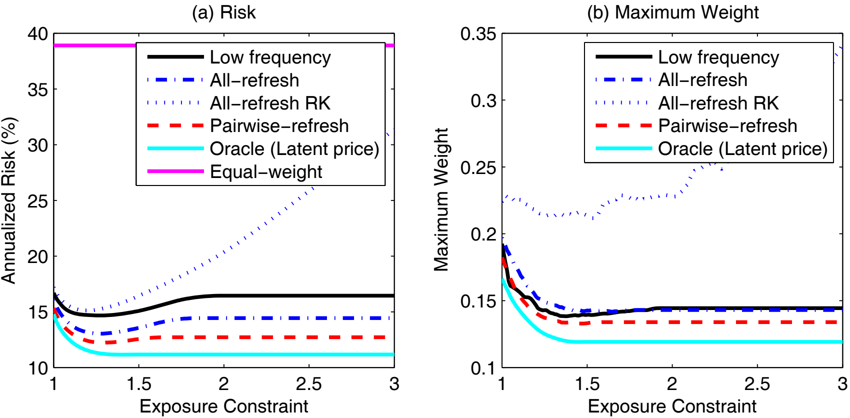
**Data**: Simulate 100 days high frequency data.

- Low-freq: past 100 days data;
- High-freq: past 10-day data

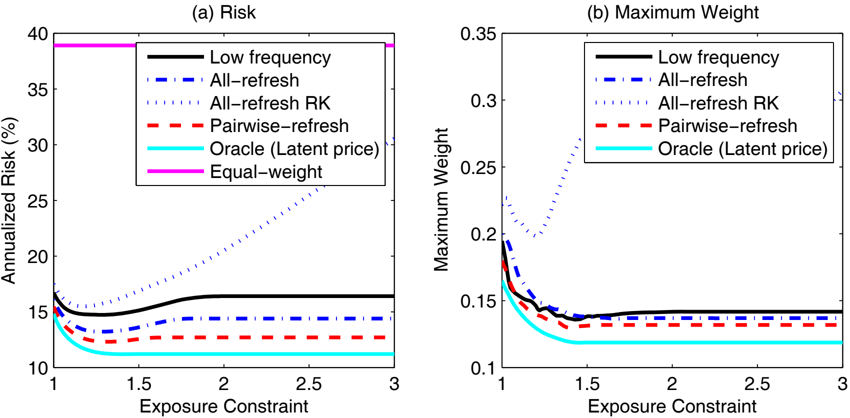
**Holding period**: holding period $\tau = 1$ or 5-days, rebalanced.

**Risk evaluation**: 15 minutes returns over 100 days (2600 returns).

# Out of sample performance ($\tau = 1$)



# Out of sample performance ($\tau = 5$)



# Conclusion

- Advocate portfolio selection with gross-exposure constraint.
- It is less sensitive to error of covariance estimation, and has little noise accumulation.
- Propose "all-fresh" and "pair-fresh" to estimates integrated covariance, derive the concentration inequalities, and demonstrate limited impact of portfolio size.
- Use of HF-data increases $n$, shortens time window, adapts to local covariation.
- Demonstrate the utility via empirical studies and simulations.

# Conclusion

- Advocate portfolio selection with gross-exposure constraint.
- It is less sensitive to error of covariance estimation, and has little noise accumulation.
- Propose "all-fresh" and "pair-fresh" to estimates integrated covariance, derive the concentration inequalities, and demonstrate limited impact of portfolio size.
- Use of HF-data increases $n$, shortens time window, adapts to local covariation.
- Demonstrate the utility via empirical studies and simulations.

# Conclusion

- Advocate portfolio selection with gross-exposure constraint.
- It is less sensitive to error of covariance estimation, and has little noise accumulation.
- Propose ”all-fresh” and ”pair-fresh” to estimates integrated covariance, derive the concentration inequalities, and demonstrate limited impact of portfolio size.
- Use of HF-data increases $n$, shortens time window, adapts to local covariation.
- Demonstrate the utility via empirical studies and simulations.

# The End

*Thank*  *You*