Portfolio Allocation using High Frequency Data

Jianqing Fan

Princeton University

With Yingying Li and Ke Yu

http://www.princeton.edu/~jqfan

September 10, 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
**Portfolio allocation**: \( \min_{w^T 1 = 1, w^T \mu = r_0} w^T \Sigma w \)

Solution: \( w = c_1 \Sigma^{-1} \mu + c_2 \Sigma^{-1} 1 \)

★ Cornerstone of modern finance.

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

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

Impact of dimensionality is large:

Risk: \( w^T \hat{\Sigma} w \).  
Allocation: \( \hat{c}_1 \hat{\Sigma}^{-1} 1 + \hat{c}_2 \hat{\Sigma}^{-1} \hat{\mu} \).
Markowitz’s Mean-variance analysis

Portfolio allocation: \( \min_{w^T1=1, w^T\mu=r_0} w^T \Sigma w \)

Solution: \( w = c_1 \Sigma^{-1} \mu + c_2 \Sigma^{-1} 1 \)

★ Cornerstone of modern finance.

★ Too sensitive on input vectors and their estimation errors.

★ More severe for large portfolios: 2000 stocks involves \( 2 \times 10^6 \) parameters! Error accumulation can be huge.

Impact of dimensionality is large:

Risk: \( w^T \hat{\Sigma} w \). Allocation: \( \hat{c}_1 \hat{\Sigma}^{-1} 1 + \hat{c}_2 \hat{\Sigma}^{-1} \hat{\mu} \).
**Portfolio allocation**: \( \min_{w^T 1 = 1, w^T \mu = r_0} w^T \Sigma w \)

Solution: \( w = c_1 \Sigma^{-1} \mu + c_2 \Sigma^{-1} 1 \)

★ 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: \( w^T \hat{\Sigma} w \).  
Allocation: \( \hat{c}_1 \hat{\Sigma}^{-1} 1 + \hat{c}_2 \hat{\Sigma}^{-1} \hat{\mu} \).
Exposure-constrained portfolio selection

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

\[
\min \quad w^T \Sigma w, \quad \|w\|_1 \leq c.
\]

\[w^T 1 = 1, \quad Aw = a\]

**Constraints**:

- expected return or sector/factor exposures via \(A\).

- **short positions**: \(w^- \leq (c - 1)/2\),

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

  \(c = 1 \implies \text{no short-sale; } c = \infty \implies \text{Markowitz problem.}\)

**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)

\[
\min \quad w^T \Sigma w, \quad \|w\|_1 \leq c.
\]

**Constraints**:

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

  \[
  w^+ + w^- \leq c, \quad w^+ - w^- = 1.
  \]

  \( c = 1 \iff \) no short-sale; \( c = \infty \iff \) 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)

\[
\begin{align*}
\min_{w} & \quad w^T \Sigma w, \\
\text{subject to} & \quad w^T 1 = 1, \quad Aw = a, \\
& \quad \|w\|_1 \leq c.
\end{align*}
\]

**Constraints**: 
- expected return or sector/factor exposures via \(A\).
- **short positions**: \(w^- \leq (c - 1)/2\), since \(w^+ + w^- \leq c\), \(w^+ - w^- = 1\).

\(c = 1 \implies\) no short-sale; \(c = \infty \implies\) Markowitz problem.

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

**Applicability**: Any coherent risk measures (Artzner et al, 1999)
Utility Approx.: Let $M(\mu, \Sigma) = w^T \mu - \lambda w^T \Sigma w$ be expected utility.

$$|M(\hat{\mu}, \hat{\Sigma}) - M(\mu, \Sigma)| \leq \|\hat{\mu} - \mu\|_\infty \|w\|_1 + \lambda |\hat{\Sigma} - \Sigma|_\infty \|w\|_1^2$$

$$\leq \|\hat{\mu} - \mu\|_\infty c + \lambda |\hat{\Sigma} - \Sigma|_\infty c^2,$$

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

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

$$|R(w, \hat{\Sigma}) - R(w, \Sigma)| \leq |\hat{\Sigma} - \Sigma|_\infty c^2,$$
**Utility Approximations**

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

\[
| M(\hat{\mu}, \hat{\Sigma}) - M(\mu, \Sigma) | \leq \| \hat{\mu} - \mu \|_\infty \| w \|_1 + \lambda \| \hat{\Sigma} - \Sigma \|_\infty \| w \|_1^2 \\
\leq \| \hat{\mu} - \mu \|_\infty c + \lambda \| \hat{\Sigma} - \Sigma \|_\infty c^2,
\]

No noise accumulation effect for moderate \( c \leq 3 \), say.

Applicable to any number of assets \( p \)

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

\[
|R(w, \hat{\Sigma}) - R(w, \Sigma)| \leq \| \hat{\Sigma} - \Sigma \|_\infty c^2,
\]
Actual and Empirical risks: \( R(w) = w^T \Sigma w, \quad R_n(w) = w^T \hat{\Sigma} w \).

Theoretical and empirical allocation vector:
\[
\begin{align*}
\mathbf{w}_{opt} &= \text{argmin}_{\|\mathbf{w}\|_1 \leq c} R(w), \quad \hat{\mathbf{w}}_{opt} = \text{argmin}_{\|\mathbf{w}\|_1 \leq c} R_n(w)
\end{align*}
\]

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
\[
\begin{align*}
|R(\hat{\mathbf{w}}_{opt}) - R(\mathbf{w}_{opt})| &\leq 2a_n c^2 \\
|R(\hat{\mathbf{w}}_{opt}) - R_n(\hat{\mathbf{w}}_{opt})| &\leq a_n c^2 \\
|R(\mathbf{w}_{opt}) - R_n(\hat{\mathbf{w}}_{opt})| &\leq a_n c^2.
\end{align*}
\]
**Actual and Empirical risks:** \( R(w) = w^T \Sigma w, \quad R_n(w) = w^T \hat{\Sigma} w. \)

Theoretical and empirical allocation vector:

\[
\begin{align*}
    w_{opt} &= \text{argmin}_{\|w\|_1 \leq c} R(w), \\
    \hat{w}_{opt} &= \text{argmin}_{\|w\|_1 \leq c} R_n(w)
\end{align*}
\]

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

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

\[
\begin{align*}
    |R(\hat{w}_{opt}) - R(w_{opt})| &\leq 2a_n c^2 \\
    |R(\hat{w}_{opt}) - R_n(\hat{w}_{opt})| &\leq a_n c^2 \\
    |R(w_{opt}) - R_n(\hat{w}_{opt})| &\leq a_n c^2.
\end{align*}
\]
Risk Approximation Theory

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

Theoretical and empirical allocation vector:

\[ w_{opt} = \text{argmin}_{\|w\|_1 \leq c} R(w), \quad \hat{w}_{opt} = \text{argmin}_{\|w\|_1 \leq c} R_n(w) \]

**Risks:** \( \sqrt{R(w_{opt})} \) — oracle, \( \sqrt{R_n(\hat{w}_{opt})} \) — empirical;

\( \sqrt{R(\hat{w}_{opt})} \) — actual risk of a selected portfolio.

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

\[
| R(\hat{w}_{opt}) - R(w_{opt}) | \leq 2a_n c^2 \\
| R(\hat{w}_{opt}) - R_n(\hat{w}_{opt}) | \leq a_n c^2 \\
| R(w_{opt}) - R_n(\hat{w}_{opt}) | \leq a_n c^2.
\]
Impact of dimensionality

Actual vs Empirical risks

Exposure Constraint $C$

Annualized Volatility(%) $(a) \ n=252, \ p=200$

Actual Risk
Empirical Risk
Optimal Risk

Exposure Constraint $C$

Annualized Volatility(%) $(b) \ n=252, \ p=500$

Actual Risk
Empirical Risk
Optimal Risk

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

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

Impact of dimensionality is limited.

Jianqing Fan
Vast portfolios & high-freq data
Portfolio Selection
with dynamic covariance
**Return and Risk** with holding period $\tau$:

\[
\text{Return} = w^T R_{t,\tau} = w^T \int_t^{t+\tau} dX_s, \quad \text{risk} = w^T \Sigma_{t,\tau} w,
\]

where $\Sigma_{t,\tau} = E_t \int_t^{t+\tau} S_u du$, allowing **stochastic** volatility and $S_u = \begin{pmatrix} \sigma_{i,j}^{(u)} \end{pmatrix}$ is instantaneous cov matrix.

**Portfolio allocation and selection:**

\[
\min_{w^T 1 = 1, \text{ Aw = a} } w^T \Sigma_{t,\tau} w, \quad \|w\|_1 \leq c.
\]
Return and Risk with holding period $\tau$:

\[
\text{Return} = w^T R_{t,\tau} = w^T \int_t^{t+\tau} dX_s, \quad \text{risk} = w^T \Sigma_{t,\tau} w,
\]

where $\Sigma_{t,\tau} = E_t \int_t^{t+\tau} S_u du$, allowing stochastic volatility and $S_u = \begin{pmatrix} \sigma_{u}^{(i,j)} \end{pmatrix}$ is instantaneous cov matrix.

Portfolio allocation and selection:

\[
\min \quad w^T \Sigma_{t,\tau} w, \quad \|w\|_1 \leq c.
\]
Prediction of Covariance Matrix

Covariance matrix is predicted based on following approximations:

**short-horizon** $\tau$:

$$\frac{1}{\tau} \sum_{t,\tau} \approx \frac{1}{h} \int_{t-h}^{t} S_u \, du$$  
(use of continuity)

**long-horizon** $\tau$:

$$\frac{1}{\tau} \sum_{t,\tau} \approx \frac{1}{h} E \int_{t-h}^{t} S_u \, du$$  
(use of ergoticity)

- Even with observed $S_u$ in the past, $\sum_{t,\tau}$ is at best approximated.
- Important to reduce the sensitivity of $w$ on the prediction of $\sum_{t,\tau}$.
- Gross-exposure constraint is an effective method.
Covariance matrix is predicted based on following approximations:

**short-horizon** \( \tau \):
\[
\frac{1}{\tau} \Sigma_{t,\tau} \approx \frac{1}{h} \int_{t-h}^{t} S_u du
\]  
(\text{use of continuity})

**long-horizon** \( \tau \):
\[
\frac{1}{\tau} \Sigma_{t,\tau} \approx \frac{1}{h} E \int_{t-h}^{t} S_u du
\]  
(\text{use of ergoticity})

- Even with observed \( S_u \) in the past, \( \Sigma_{t,\tau} \) is at best **approximated**.
- Important to reduce the **sensitivity** of \( w \) 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:** $Y_{ti} = X_{ti} + \varepsilon_{ti}$, \quad $X_{ti}$ — latent log-price, $\eta^2 = \text{var}(\varepsilon)$

- 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: Use once every $K$ points

$$RV_{K,i} = \sum_{j=1}^{n_s} (Y_{t_i+jK} - Y_{t_i+(j-1)K})^2, \quad n_s = n/K, \quad \Theta = \int_{t-h}^{t} \sigma^2 u du.$$ 

$$= \Theta + 2n_s \eta^2 + \left[4n_s E \varepsilon^4 + \frac{2}{n_s} \int \sigma^4_t dt \right]^{1/2} \cdot N(0, 1),$$

Averaging: $[Y]^{(K)} = \frac{1}{K} \sum_{i=0}^{K-1} R_{K,i} = \frac{1}{K} \sum_{i=1}^{n-K} (Y_{t_i+K} - Y_{t_i})^2$

$$\approx \Theta + 2n_s \eta^2 + \left[\frac{4n_s}{K} E \varepsilon^4 + \frac{4}{3n_s} \int \sigma^4_t dt \right]^{1/2} \cdot N(0, 1).$$
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 \(O\left(\frac{(\log p)^{1/2}}{\eta_{\min}^{1/6}}\right)\).
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 \( O\left( \frac{(\log p)^{1/2}}{n^{1/6}} \right) \).
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):

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

where \(\tilde{n}\) is no. of synchronized data, and

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

2 Realized Covariance (BNHLS, 08): log-return \(y_t\)

\[
K(X) = \sum_{h=-H}^{H} k \left( \frac{h}{H+1} \right) \Gamma_h, \quad \Gamma(h) = \sum_{j=|h|+1}^{n} y_j y_{j-|h|}'
\]

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

\[
\langle \hat{Y}_1, \hat{Y}_2 \rangle = \frac{1}{4} \left\{ \langle \hat{Y}_1 + \hat{Y}_2, \hat{Y}_1 + \hat{Y}_2 \rangle_{QMLE} - \langle \hat{Y}_1 - \hat{Y}_2, \hat{Y}_1 - \hat{Y}_2 \rangle_{QMLE} \right\}
\]
Two-Scale Realized Covariance (Zhang, 09):

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

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

$$[Y_1, Y_2]^{(K)} = \frac{1}{K} \sum_{i=K}^{\tilde{n}} (Y_{1,ti} - Y_{1,ti-K})(Y_{2,ti} - Y_{2,ti-K}),$$

subsam cov

Realized Covariance (BNHLS, 08): log-return $y_t$

$$K(X) = \sum_{h=-H}^{H} k \left( \frac{h}{H+1} \right) \Gamma_h, \quad \Gamma(h) = \sum_{j=|h|+1}^{n} y_j y_{j-|h|}'$$

QMLE (Aït-Sahalia, Fan and Xiu, 2010)

$$\langle \hat{Y}_1, \hat{Y}_2 \rangle = \frac{1}{4} \left\{ \langle \hat{Y}_1 + \hat{Y}_2, \hat{Y}_1 + \hat{Y}_2 \rangle_{QMLE} - \langle \hat{Y}_1 - \hat{Y}_2, \hat{Y}_1 - \hat{Y}_2 \rangle_{QMLE} \right\}$$
Estimation of integrated covariance

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

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

where \( \tilde{n} \) is no. of synchronized data, and

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

2. Realized Covariance (BNHLS, 08): log-return \( y_t \)

\[
K(X) = \sum_{h=-H}^{H} k\left(\frac{h}{H+1}\right) \Gamma_h, \quad \Gamma(h) = \sum_{j=|h|+1}^{n} y_j y'_{j-|h|}
\]

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

\[
\langle \widehat{Y_1, Y_2} \rangle = \frac{1}{4} \{ \langle \widehat{Y_1 + Y_2, Y_1 + Y_2} \rangle_{QMLE} - \langle \widehat{Y_1 - Y_2, Y_1 - Y_2} \rangle_{QMLE} \}
\]
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 \( \text{cor}(B_t^{(i)}, B_t^{(j)}) = \rho_t^{(i,j)} \).
2. **Volatility:** \( |\sigma_t^{(i)}| < C_\sigma \).
3. **Refresh time:** \( \sup_j |\tau_j - \tau_{j-1}| \leq C_\tau / n_1 \).
4. **Noise:** \( \{\epsilon_t^{Y_i}\} \) are independent, also independent of \( X^{(i)} \).
Theorem 4. For large $x$ that satisfies $|x| \leq c\tilde{n}^{1/6}$,

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

Conditions:

1. **Log-price**: $dX_t^{(i)} = \sigma_t^{(i)} dB_t^{(i)}$ with $\text{cor}(B_t^{(i)}, B_t^{(j)}) = \rho_t^{(i,j)}$.

2. **Volatility**: $|\sigma_t^{(i)}| < C_\sigma$.

3. **Refresh time**: $\sup_j |\tau_j - \tau_{j-1}| \leq C / n_1$

4. **Noise**: $\{\varepsilon_t^{Y_i}\}$ are independent, also independent of $X^{(i)}$. 

Jianqing Fan  
Vast portfolios & high-freq data
Applications to Portfolio Allocation
**Portfolio Optimization**

**Portfolio allocation**: \( \min_{w^T1 = 1, \|w\|_1 \leq c} w^T \hat{\Sigma} w \). 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:

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

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**: \( \min_{\mathbf{w}^T \mathbf{1} = 1, \| \mathbf{w} \|_1 \leq c} \mathbf{w}^T \hat{\Sigma} \mathbf{w} \). 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:

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

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: \[ \min_{\mathbf{w}^T \mathbf{1} = 1, \|\mathbf{w}\|_1 \leq c} \mathbf{w}^T \hat{\Sigma} \mathbf{w}. \] 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:

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

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

\[
|\hat{\Sigma} - \Sigma|_\infty = O\left(\frac{\sqrt{\log p}}{n_*^{1/6}}\right).
\]
Need of projection: Programming algorithms require $\Sigma \geq 0$.

**Projection 1:** $A_1^+ = \Gamma^T \text{diag}(\lambda_1^+, \cdots, \lambda_n^+) \Gamma$, for a symmetric matrix with SVD $A = \Gamma^T \text{diag}(\lambda_1, \cdots, \lambda_n) \Gamma$.

**Projection 2:** $A_2^+ = (A - \lambda_{\min} I) / (1 - \lambda_{\min})$, 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.
Need of projection: Programming algorithms require $\Sigma \geq 0$.

Projection 1: $A_1^+ = \Gamma^T \text{diag}(\lambda_1^+, \cdots, \lambda_n^+) \Gamma$, for a symmetric matrix with SVD $A = \Gamma^T \text{diag}(\lambda_1, \cdots, \lambda_n) \Gamma$.

Projection 2: $A_2^+ = (A - \lambda_{\min} I_p)/(1 - \lambda_{\min})$, 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 $\hat{\Sigma} \geq 0$.

**Projection 1**: $A_1^+ = \Gamma^T \text{diag}(\lambda_1^+, \cdots, \lambda_n^+) \Gamma$, for a symmetric matrix with SVD $A = \Gamma^T \text{diag}(\lambda_1, \cdots, \lambda_n) \Gamma$.

**Projection 2**: $A_2^+ = (A - \lambda_{\text{min}} I_p)/(1 - \lambda_{\text{min}})$, where $\lambda_{\text{min}}$ 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.
It appears projections **distort** more “pairwise refresh” method than “all refresh”. Thus, the smaller componentwise estimation errors might not be materialized in implementation.

Risk approximation is an upper bound, not necessarily tight.

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

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 $\hat{\Sigma}$ 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: 207,630,360. 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 $26 \times 90 = 2340$ returns), excluding overnight holding risks.

- High frequency $h = 10$ days; low frequency $h = 100$ days
Summary of Trading Frequencies

<table>
<thead>
<tr>
<th>Max No of Trades</th>
<th>Min No of Trades</th>
<th>Median No of Trades</th>
</tr>
</thead>
<tbody>
<tr>
<td>2\times10^5</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>6</td>
<td>10</td>
<td></td>
</tr>
</tbody>
</table>

Jianqing Fan
Vast portfolios & high-freq data
An empirical result ($\tau = 1$)

(a) Risk

Exposure Constraint

Annualized Risk (%)

Low frequency
All-refresh
All-refresh RK
Pairwise-refresh
Equal-weight

(b) Maximum Weight

Exposure Constraint

Low frequency
All-refresh
All-refresh RK
Pairwise-refresh
An empirical result ($\tau = 5$)

(a) Risk

(b) Maximum Weight

- Low frequency
- All-refresh
- All-refresh RK
- Pairwise-refresh
- Equal-weight

Exposure Constraint

Annualized Risk (%)

Maximum Weight

Jianqing Fan  Vast portfolios & high-freq data
A Simulation Study
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):

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

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

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

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

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

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

$$(\mu^{(i)}, \beta_0^{(i)}, \beta_1^{(i)}, \alpha^{(i)}) = (0.03, -1, 0.75, 1/40) \otimes U(i),$$

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

Vast portfolios & high-freq data
Stochastic models

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

\[
dX^{(i)}_t = \mu^{(i)} dt + \rho^{(i)} \sigma^{(i)}_t dB^{(i)}_t + \sqrt{1 - (\rho^{(i)})^2} \sigma^{(i)}_t dW_t + \lambda^{(i)} dZ^{(i)}_t,
\]

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

**Stochastic volatility**: \( \eta^{(i)}_t = \log \sigma^{(i)}_t \) follows Vasicek model (OU):

\[
d\eta^{(i)}_t = \alpha^{(i)} (\beta^{(i)}_0 - \eta^{(i)}_t) dt + \beta^{(1)}_1 dB^{(i)}_t.
\]

**Choice of parameter**: \( \rho^{(i)} = -0.7, \lambda^{(i)} = \exp(\beta^{(i)}_0), \)

\[
(\mu^{(i)}, \beta^{(i)}_0, \beta^{(i)}_1, \alpha^{(i)}) = (0.03, -1, .75, 1/40) \otimes U^{(i)},
\]

where \( U^{(i)} \sim_{i.i.d.} \text{Unif}(0.7, 1.3)^\otimes 4 \).
**Trading frequency**: Poisson process with $\lambda_i = 0.02i \times 23400$ — no. of seconds / day.

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

**Microstructural noise**: $Y_{tij}^{(i)} = X_{tij}^{(i)} + N(0, 0.0005^2)$. 
Examples of realized volatilities and prices

- Varying volatility, but relatively calm.
Risk approximation: In-sample evaluation

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

\[
\begin{align*}
  w_3 &= (1 + 2/p, -1, 1/p, \cdots, 1/p)^T, \\
  w_3 &= (2, -1, 0, \cdots, 0)^T
\end{align*}
\]

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

<table>
<thead>
<tr>
<th>Portfolio</th>
<th>Latent Median(RSD)</th>
<th>All-refresh TSRV Median(RSD)</th>
<th>All-refresh RK Median(RSD)</th>
<th>Pairwise TSRV Median(RSD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( w_1 )</td>
<td>0.440 (0.0032)</td>
<td>0.387 (0.107)</td>
<td>0.434 (0.024)</td>
<td>0.419 (0.069)</td>
</tr>
<tr>
<td>( w_2 )</td>
<td>0.591 (0.0060)</td>
<td>0.522 (0.125)</td>
<td>0.623 (0.025)</td>
<td>0.593 (0.128)</td>
</tr>
<tr>
<td>( w_3 )</td>
<td>0.539 (0.0044)</td>
<td>0.469 (0.090)</td>
<td>0.583 (0.025)</td>
<td>0.520 (0.073)</td>
</tr>
<tr>
<td>( w_4 )</td>
<td>0.844 (0.0077)</td>
<td>0.753 (0.174)</td>
<td>0.922 (0.041)</td>
<td>0.839 (0.178)</td>
</tr>
</tbody>
</table>
Risk approximation: In-sample evaluation

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

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

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

<table>
<thead>
<tr>
<th>Portfolio</th>
<th>Latent Median(RSD)</th>
<th>All-refresh Median(RSD)</th>
<th>All-refresh TSRV</th>
<th>All-refresh RK Median(RSD)</th>
<th>Pairwise TSRV Median(RSD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$w_1$</td>
<td><strong>0.440</strong> (0.0032)</td>
<td>0.387 (0.107)</td>
<td><strong>0.434</strong> (0.024)</td>
<td><strong>0.419</strong> (0.069)</td>
<td></td>
</tr>
<tr>
<td>$w_2$</td>
<td>0.591 (0.0060)</td>
<td>0.522 (0.125)</td>
<td>0.623 (0.025)</td>
<td>0.593 (0.128)</td>
<td></td>
</tr>
<tr>
<td>$w_3$</td>
<td>0.539 (0.0044)</td>
<td>0.469 (0.090)</td>
<td>0.583 (0.025)</td>
<td>0.520 (0.073)</td>
<td></td>
</tr>
<tr>
<td>$w_4$</td>
<td>0.844 (0.0077)</td>
<td>0.753 (0.174)</td>
<td>0.922 (0.041)</td>
<td>0.839 (0.178)</td>
<td></td>
</tr>
</tbody>
</table>
### Median and RSD of Risk Estimation Errors

<table>
<thead>
<tr>
<th>Portfolio</th>
<th>All-refresh TSRV Median(RSD)</th>
<th>All-refresh RK Median(RSD)</th>
<th>Pairwise TSRV Median(RSD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$w_1$</td>
<td><strong>0.0889</strong> (0.0769)</td>
<td>0.0183 (0.0153)</td>
<td><strong>0.0547</strong> (0.0439)</td>
</tr>
<tr>
<td>$w_2$</td>
<td>0.1054 (0.0700)</td>
<td>0.0344 (0.0272)</td>
<td>0.0804 (0.0813)</td>
</tr>
<tr>
<td>$w_3$</td>
<td>0.0936 (0.0665)</td>
<td>0.0437 (0.0300)</td>
<td>0.0599 (0.0593)</td>
</tr>
<tr>
<td>$w_4$</td>
<td>0.1470 (0.1022)</td>
<td>0.0794 (0.0393)</td>
<td>0.1089 (0.0941)</td>
</tr>
</tbody>
</table>

### Median and RSD of $a_p$

<table>
<thead>
<tr>
<th>All-refresh TSRV Median(RSD)</th>
<th>All-refresh RK Median(RSD)</th>
<th>Pairwise TSRV Median(RSD)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>0.2476</strong> (0.1460)</td>
<td>0.0603 (0.0270)</td>
<td><strong>0.1730</strong> (0.0746)</td>
</tr>
</tbody>
</table>
Evaluation of portfolio allocation: In-sample risk ($\tau = 1$)

In-the-sample Risk Comparison for Optimal Allocation

- Oracle (Latent price)
- All-refresh
- All-refresh RK
- Pairwise-refresh

Jianqing Fan
Vast portfolios & high-freq data
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 ($\tau = 1$)

(a) Risk

(b) Maximum Weight

- Low frequency
- All-refresh
- All-refresh RK
- Pairwise-refresh
- Oracle (Latent price)
- Equal-weight

Annualized Risk (%)

Maximum Weight
Out of sample performance ($\tau = 5$)

(a) Risk

- Low frequency
- All-refresh
- All-refresh RK
- Pairwise-refresh
- Oracle (Latent price)
- Equal-weight

(b) Maximum Weight

- Low frequency
- All-refresh
- All-refresh RK
- Pairwise-refresh
- Oracle (Latent price)
Conclusion

- Advocate portfolio selection with gross-exposure constraint. Less sensitive to estimation errors, & little noise accumulation.

- Propose "all-fresh" and "pair-fresh" methods, 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.
Advocate portfolio selection with gross-exposure constraint. Less sensitive to estimation errors, & little noise accumulation.

Propose "all-fresh" and "pair-fresh" methods, 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.
Advocate portfolio selection with gross-exposure constraint. Less sensitive to estimation errors, & little noise accumulation.

Propose "all-fresh" and "pair-fresh" methods, 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.
Advocate portfolio selection with gross-exposure constraint. Less sensitive to estimation errors, & little noise accumulation.

Propose ”all-fresh” and ”pair-fresh” methods, 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