Condition Number Regularized Covariance Estimation

Supplemental Materials

Joong-Ho Won *School of Industrial Management Engineering, Korea University, Seoul, Korea*

Johan Lim *Department of Statistics, Seoul National University, Seoul, Korea*

Seung-Jean Kim *Citi Capital Advisors, New York, NY, USA*

Bala Rajaratnam *Department of Statistics, Stanford University, Stanford, CA, USA*

Abstract

This document contains additional proofs of Theorems 2 and 3, and Proposition 2; additional figures illustrating Bayesian prior densities; additional figure illustrating risk simulations; details of the empirical minimum variance rebalancing study.

A Additional proofs

Proof of Theorem 2. Suppose the spectral decomposition of the k-th fold covariance matrix estimate $\widehat{\Sigma}_\nu^{[-k]}$, with $\kappa_{\max} = \nu$, is given by

$$\widehat{\Sigma}_\nu^{[-k]} = Q^{[-k]} \mathrm{diag}\big(\widehat{\lambda}_1^{[-k]}, \ldots, \widehat{\lambda}_p^{[-k]}\big)\big(Q^{[-k]}\big)^T$$

with

$$\widehat{\lambda}_i^{[-k]} = \begin{cases} v^{[-k]*} & \text{if } l_i^{[-k]} < v^{[-k]*} \\ l_i^{[-k]} & \text{if } v^{[-k]*} \leq l_i^{[-k]} < \nu v^{[-k]*} \\ \nu v^{[-k]*} & \text{if } l_i^{[-k]} \geq \nu v^{[-k]*}, \end{cases}$$

where $l_i^{[-k]}$ is the i-th largest eigenvalue of the k-th fold sample covariance matrix $S^{[-k]}$, and $v^{[-k]*}$ is obtained by the method described in Section 2. Since $\widehat{\Sigma}_\nu^{[-k]} = S^{[-k]}$ if $\nu \geq l_1^{[-k]}/l_p^{[-k]} = \mathrm{cond}(S^{[-k]})$,

$$\widehat{\kappa}_{\max} \leq \max_{k=1,\ldots,K} \mathrm{cond}(S^{[-k]}). \tag{27}$$

The right hand side of (27) converges in probability to the condition number κ of the true covariance matrix, as n increases while p is fixed. Hence,

$$\lim_{n\to\infty} P\Big(\widehat{\kappa}_{\max} \leq \kappa\Big) = 1.$$

We now show that

$$\lim_{n\to\infty} P\Big(\widehat{\kappa}_{\max} \geq \kappa\Big) = 1.$$

by showing that $\widehat{\mathrm{PR}}(\nu)$ is an asymptotically decreasing function in ν.

Recall that

$$\widehat{\mathrm{PR}}(\nu) = -\frac{1}{n}\sum_{k=1}^{K} l_k\big(\widehat{\Sigma}_\nu^{[-k]}, X_k\big),$$

where

$$l_k\big(\widehat{\Sigma}_\nu^{[-k]}, X_k\big) \quad = \quad -(n_k/2)\Big[\mathbf{tr}\left\{\big(\widehat{\Sigma}_\nu^{[-k]}\big)^{-1} X_k X_k^T/n_k\right\} - \log\det\big(\widehat{\Sigma}_\nu^{[-k]}\big)^{-1}\Big],$$

which, by the definition of $\widehat{\Sigma}_\nu^{[-k]}$, is everywhere differentiable but at a finite number of points.

To see the asymptotic monotonicity of $\widehat{\mathrm{PR}}(\nu)$, consider the derivative $-\partial l_k\big(\widehat{\Sigma}_\nu^{[-k]}, X_k\big)/\partial\nu$:

$$
\begin{aligned}
-\frac{\partial l_k\big(\widehat{\Sigma}_\nu^{[-k]}, X_k\big)}{\partial \nu} &= \frac{n_k}{2}\left[\mathbf{tr}\left(\big(\widehat{\Sigma}_\nu^{[-k]}\big)^{-1}\frac{\partial \widehat{\Sigma}_\nu^{[-k]}}{\partial \nu}\right) + \mathbf{tr}\left\{\frac{\partial \big(\widehat{\Sigma}_\nu^{[-k]}\big)^{-1}}{\partial \nu}\Big(X_k X_k^T/n_k\Big)\right\}\right] \\
&= \frac{n_k}{2}\left[\mathbf{tr}\left(\big(\widehat{\Sigma}_\nu^{[-k]}\big)^{-1}\frac{\partial \widehat{\Sigma}_\nu^{[-k]}}{\partial \nu}\right) + \mathbf{tr}\left\{\frac{\partial \big(\widehat{\Sigma}_\nu^{[-k]}\big)^{-1}}{\partial \nu}\widehat{\Sigma}_\nu^{[-k]}\right\}\right. \\
&\quad \left. + \mathbf{tr}\left\{\frac{\partial \big(\widehat{\Sigma}_\nu^{[-k]}\big)^{-1}}{\partial \nu}\Big(X_k X_k^T/n_k - \widehat{\Sigma}_\nu^{[-k]}\Big)\right\}\right] \\
&= \frac{n_k}{2}\left[\frac{\partial}{\partial \nu}\,\mathbf{tr}\left(\big(\widehat{\Sigma}_\nu^{[-k]}\big)^{-1}\widehat{\Sigma}_\nu^{[-k]}\right) + \mathbf{tr}\left\{\frac{\partial \big(\widehat{\Sigma}_\nu^{[-k]}\big)^{-1}}{\partial \nu}\Big(X_k X_k^T/n_k - \widehat{\Sigma}_\nu^{[-k]}\Big)\right\}\right] \\
&= \frac{n_k}{2}\,\mathbf{tr}\left\{\frac{\partial \tilde{\Sigma}_\nu^{-1}}{\partial \nu}\Big(X_k X_k^T/n_k - \widehat{\Sigma}_\nu^{[-k]}\Big)\right\}.
\end{aligned}
$$

As n (hence n_k) increases, $\widehat{\Sigma}_\nu^{[-k]}$ converges almost surely to the inverse of the solution to the following optimization problem

$$
\begin{array}{ll}
\text{minimize} & \mathbf{tr}(\Omega\Sigma) - \log\det\Omega \\
\text{subject to} & \mathrm{cond}(\Omega) \le \nu,
\end{array}
$$

with Σ and ν replacing S and $\kappa_{\max}$ in (8). We denote the limit of $\widehat{\Sigma}_\nu^{[-k]}$ by $\tilde{\Sigma}_\nu$. For the spectral decomposition of Σ

$$
\Sigma = R\ \mathrm{diag}\big(\lambda_1, \ldots, \lambda_p\big) R^T, \tag{28}
$$

$\tilde{\Sigma}_\nu$ is given as

$$
\tilde{\Sigma}_\nu = R\ \mathrm{diag}\big(\psi_1(\nu), \ldots, \psi_p(\nu)\big)\ R^T, \tag{29}
$$

where, for some $\tau(\nu) > 0$,

$$
\psi_i(\nu) = \begin{cases} \tau(\nu) & \text{if} \quad \lambda_i \le \tau(\nu) \\ \lambda_i & \text{if} \quad \tau(\nu) < \lambda_i \le \nu\tau(\nu) \\ \nu\tau(\nu), & \text{if} \quad \nu\tau(\nu) < \lambda_i. \end{cases}
$$

Recall from Proposition 1 that $\tau(\nu)$ is decreasing in ν and $\nu\tau(\nu)$ is increasing.

Let c_k be the limit of $n_k/(2n)$ when both n and n_k increases. Then, $X_k X_k^T/n_k$ converges almost surely to Σ. Thus,

$$
-\frac{1}{n}\frac{\partial l_k\big(\widehat{\Sigma}_\nu^{[-k]}, X_k\big)}{\partial \nu} \to c_k\,\mathbf{tr}\left\{\frac{\partial \tilde{\Sigma}_\nu^{-1}}{\partial \nu}\Big(\Sigma - \tilde{\Sigma}_\nu\Big)\right\}, \qquad \text{almost surely.} \tag{30}
$$

If $\nu \ge \kappa$, then $\tilde{\Sigma}_\nu = \Sigma$, and the RHS of (30) degenerates to 0. Now consider the case that $\nu < \kappa$. From (29),

$$
\frac{\partial \tilde{\Sigma}_\nu^{-1}}{\partial \nu} = R\frac{\partial \Psi^{-1}}{\partial \nu}R^T = R\ \mathrm{diag}\left(\frac{\partial \psi_1^{-1}}{\partial \nu}, \ldots, \frac{\partial \psi_p^{-1}}{\partial \nu}\right)R^T,
$$

where

$$\frac{\partial \psi_i^{-1}}{\partial \nu} = \begin{cases} -\frac{1}{\tau(\nu)^2}\frac{\partial \tau(\nu)}{\partial \nu} & (\geq 0) \quad \text{if} \quad \lambda_i \leq \tau(\nu) \\ 0 & \qquad\quad \text{if} \quad \tau(\nu) < \lambda_i \leq \nu\tau(\nu) \\ -\frac{1}{\nu^2\tau(\nu)^2}\frac{\partial(\nu\tau(\nu))}{\partial \nu} & (\leq 0) \quad \text{if} \quad \nu\tau(\nu) < \lambda_i. \end{cases}$$

From (28) and (29),

$$\Sigma - \tilde{\Sigma}_\nu = R\text{diag}\big(\lambda_1 - \psi_1, \ldots, \lambda_p - \psi_p\big)R^T,$$

where

$$\lambda_i - \psi_i = \begin{cases} \lambda_i - \tau(\nu) & (\leq 0) \quad \text{if} \quad \lambda_i \leq \tau(\nu) \\ 0 & \qquad\quad \text{if} \quad \lambda_i \leq \nu\tau(\nu) \\ \lambda_i - \nu\tau(\nu) & (\geq 0) \quad \text{if} \quad \nu\tau(\nu) < \lambda_i. \end{cases}$$

Therefore,

$$\mathbf{tr}\left\{\frac{\partial \tilde{\Sigma}_\nu^{-1}}{\partial \nu}\Big(\Sigma - \tilde{\Sigma}_\nu\Big)\right\} = \sum_i \frac{\partial \psi_i^{-1}}{\partial \nu} \cdot (\lambda_i - \psi_i) \leq 0.$$

In other words, the RHS of (30) is less than 0 and the almost sure limit of $\widehat{\text{PR}}(\nu)$ is decreasing in ν.

By definition, $\widehat{\text{PR}}(\widehat{\kappa}_{\max}) \leq \widehat{\text{PR}}(\kappa)$. From this and the asymptotic monotonicity of $\widehat{\text{PR}}(\nu)$, we conclude that

$$\lim_{n\to\infty} \mathbf{pr}\left(\widehat{\kappa}_{\max} \geq \kappa\right) = 1.$$

□

Proof of Proposition 2. We are given that

$$\pi(\lambda_1, \ldots, \lambda_p) = \exp\left(-g_{\max}\frac{\lambda_1}{\lambda_p}\right) \qquad \lambda_1 \geq \cdots \geq \lambda_p > 0.$$

Now

$$\int_C \pi(\lambda_1, \ldots, \lambda_p)d\underline{\lambda} = \int_C \exp\left(-g_{\max}\frac{\lambda_1}{\lambda_p}\right)d\underline{\lambda},$$

where $C = \{\lambda_1 \geq \cdots \geq \lambda_p > 0\}$.

Let us now make the following change of variables: $x_i = \lambda_i - \lambda_{i+1}$ for $i = 1, 2, .., p-1$, and $x_p = \lambda_p$. The inverse transformation yields $\lambda_i = \sum_{j=i}^p x_j$ for $i = 1, 2, .., p$. It is straightforward to verify that the Jacobian of this transformation is given by $|J| = 1$.

Now we can therefore rewrite the integral above as

$$\begin{aligned}
\int_C \exp\left(-g_{\max}\frac{\lambda_1}{\lambda_p}\right) d\underline{\lambda} &= \int_{\mathbb{R}_1^p} \exp\left(-g_{\max}\frac{x_1+\cdots+x_p}{x_p}\right) dx_1\cdots dx_p \\
&= \mathrm{e}^{-g_{\max}} \int \left[\prod_{i=1}^{p-1}\int \exp\left(-g_{\max}\frac{x_i}{x_p}\right) dx_i\right] dx_p \\
&= \mathrm{e}^{-g_{\max}} \int_0^\infty \left(\frac{x_p}{g_{\max}}\right)^{p-1} dx_p \\
&= \frac{\mathrm{e}^{-g_{\max}}}{g_{\max}^{p-1}} \int_0^\infty x_p^{p-1} dx_p \\
&= \infty.
\end{aligned}$$

To prove that the posterior yields a proper distribution we proceed as follows:

$$\begin{aligned}
&\int_C \pi(\underline{\lambda}) f(\underline{\lambda}, \underline{l}) d\underline{\lambda} \\
\propto\ & \int_C \exp\left(-\frac{n}{2}\sum_{i=1}^p \frac{l_i}{\lambda_i}\right)\left(\prod_{i=1}^p \lambda_i\right)^{-\frac{n}{2}} \exp\left(-g_{\max}\frac{\lambda_1}{\lambda_p}\right) d\underline{\lambda} \\
\leq\ & \int_C \exp\left(-\frac{n}{2}\sum_{i=1}^p \frac{l_p}{\lambda_i}\right)\left(\prod_{i=1}^p \lambda_i\right)^{-\frac{n}{2}} \exp\left(-g_{\max}\frac{\lambda_1}{\lambda_p}\right) d\underline{\lambda} \ \text{ as } \ l_p \leq l_i \ \ \forall i = 1,\ldots,p \\
\leq\ & \int_C \exp\left(-\frac{n}{2}\sum_{i=1}^p \frac{l_p}{\lambda_i}\right)\left(\prod_{i=1}^p \lambda_i\right)^{-\frac{n}{2}} \mathrm{e}^{-g_{\max}} d\underline{\lambda} \ \text{ as } \ \frac{\lambda_1}{\lambda_p} \geq 1 \\
\leq\ & \mathrm{e}^{-g_{\max}} \prod_{i=1}^p \left(\int_0^\infty \exp\left(-\frac{n}{2}\frac{l_p}{\lambda_i}\right)\lambda_i^{-\frac{n}{2}} d\lambda_i\right).
\end{aligned}$$

The above integrand is the density of the inverse gamma distribution and therefore the corresponding integral above has a finite normalizing constant and thus yielding a proper posterior.

□

Proof of Theorem 3. (i) The conditional risk of $\widetilde{\Sigma}(\kappa_{\max},\omega)$, given the sample eigenvalues $\underline{l} = (l_1,\ldots,l_p)$, is

$$\mathbf{E}\left(\mathcal{L}_{\text{ent}}(\widetilde{\Sigma}(\kappa_{\max},\omega),\Sigma)\big|\underline{l}\right) = \sum_{i=1}^p \left\{\tilde{\lambda}_i \,\mathbf{E}\left(a_{ii}(Q)\big|\underline{l}\right) - \log\tilde{\lambda}_i\right\} + \log\det\Sigma - p,$$

where $a_{ii}(Q) = \sum_{j=1}^p q_{ji}^2\lambda_j^{-1}$ and q_{ji} is the (j,i)-th element of the orthogonal matrix Q. This is because

$$\begin{aligned}
\mathcal{L}_{\text{ent}}\left(\tilde{\Sigma}(\kappa_{\max},\omega),\Sigma\right) &= \mathbf{tr}\left(\tilde{\Lambda}A(Q)\right) - \log\det\tilde{\Lambda} + \log\det\Sigma - p \\
&= \sum_{i=1}^p \left\{\tilde{\lambda}_i a_{ii}(Q) - \log\tilde{\lambda}_i\right\} + \log\det\Sigma - p, \qquad (31)
\end{aligned}$$

where $A(Q) = Q^T \Sigma^{-1} Q$.

In (31), the summand has the form

$$x \, \mathbf{E}\left(a_{ii}(Q) \big| \underline{l}\right) - \log x$$

whose minimum is achieved at $x = 1/\mathbf{E}\left(a_{ii}(Q) \big| \underline{l}\right)$. Since $\sum_{j=1}^{p} q_{ji}^2 = 1$ and $\Sigma^{-1} \in \mathcal{D}\big(\kappa_{\max}, 1/(\kappa_{\max}\omega)\big)$ if and only if $\Sigma \in \mathcal{D}\big(\kappa_{\max}, \omega\big)$, we have $1/(\kappa_{\max}\omega) \leq a_{ii}(Q) \leq 1/\omega$. Hence $1/\mathbf{E}\left(a_{ii}(Q) \big| \underline{l}\right)$ lies between ω and $\kappa_{\max}\omega$ almost surely. Therefore,

1. If $l_i \leq \omega$, then $\tilde{\lambda}_i = \omega$ and

$$\tilde{\lambda}_i \, \mathbf{E}\left(a_{ii}(Q) \big| \underline{l}\right) - \log \tilde{\lambda}_i \leq l_i \, \mathbf{E}\left(a_{ii}(Q) \big| \underline{l}\right) - \log l_i.$$

2. If $\omega \leq l_i < \kappa_{\max}\omega$, then $\tilde{\lambda}_i = l_i$ and

$$\tilde{\lambda}_i \, \mathbf{E}\left(a_{ii}(Q) \big| \underline{l}\right) - \log \tilde{\lambda}_i = l_i \, \mathbf{E}\left(a_{ii}(Q) \big| \underline{l}\right) - \log l_i.$$

3. If $l_i \geq \kappa_{\max}\omega$, then $\tilde{\lambda}_i = \kappa_{\max}\omega$ and

$$\tilde{\lambda}_i \, \mathbf{E}\left(a_{ii}(Q) \big| \underline{l}\right) - \log \tilde{\lambda}_i \leq l_i \, \mathbf{E}\left(a_{ii}(Q) \big| \underline{l}\right) - \log l_i.$$

Thus,

$$\sum_{i=1}^{p} \left\{ \tilde{\lambda}_i \, \mathbf{E}\left(a_{ii}(Q) \big| \underline{l}\right) - \log \tilde{\lambda}_i \right\} \leq \sum_{i=1}^{p} \left\{ l_i \, \mathbf{E}\left(a_{ii}(Q) \big| \underline{l}\right) - \log l_i \right\}$$

and the risk with respect to the entropy loss is

$$\begin{aligned} \mathcal{R}_{\text{ent}}\big(\widetilde{\Sigma}(\kappa_{\max}, \omega)\big) &= \mathbf{E}\left[\sum_{i=1}^{p} \left\{ \tilde{\lambda}_i \, \mathbf{E}\left(a_{ii}(Q) \big| \underline{l}\right) - \log \tilde{\lambda}_i \right\} \right] \\ &\leq \mathbf{E}\left[\sum_{i=1}^{p} \left\{ l_i \, \mathbf{E}\left(a_{ii}(Q) \big| \underline{l}\right) - \log l_i \right\} \right] \\ &= \mathcal{R}_{\text{ent}}\big(S\big). \end{aligned}$$

In other words, $\widetilde{\Sigma}(\kappa_{\max}, \omega)$ has a smaller risk than S, provided $\underline{\lambda}^{-1} \in \mathcal{D}(\kappa_{\max}, \omega)$.

(ii) Suppose the true covariance matrix Σ has the spectral decomposition $\Sigma = R \Lambda R^T$ with R orthogonal and $\Lambda = \text{diag}\big(\lambda_1, \ldots, \lambda_p\big)$. Let $A = R \Lambda^{1/2}$, then $S_0 \triangleq A^T S A$ has the same distribution as the sample covariance matrix observed from a p-variate Gaussian distribution with the identity covariance matrix. From the variational definition of the largest eigenvalue l_1 of S, we obtain

$$l_1 = \max_{v \neq 0} \frac{v^T S v}{v^T v} = \max_{w \neq 0} \frac{w^T S_0 w}{w^T \Lambda^{-1} w},$$

where $w = A^T v$. Furthermore, since for any $w \neq 0$,

$$\lambda_1^{-1} = \min_{w \neq 0} \frac{w^T \Lambda^{-1} w}{w^T w} \leq \frac{w^T \Lambda^{-1} w}{w^T w},$$

we have

$$l_1 \le \lambda_1 \max_{w \neq 0} \frac{w^T S_0 w}{w^T w} = \lambda_1 e_1, \tag{32}$$

where e_1 is the largest eigenvalue of S_0. Using essentially the same argument, we can show that

$$l_p \ge \lambda_p e_p, \tag{33}$$

where e_p is the smallest eigenvalue of S_0. Then, from the results by Geman (1980) and Silverstein (1985), we see that

$$\mathbf{pr}\left(\left\{e_1 \le \left(1+\sqrt{\gamma}\right)^2,\ e_p \ge \left(1-\sqrt{\gamma}\right)^2\right\} \text{ eventually}\right) = 1. \tag{34}$$

The combination of (32)–(34) leads to

$$\mathbf{pr}\left(\left\{l_1 \le \lambda_1\left(1+\sqrt{\gamma}\right)^2,\ l_p \ge \lambda_p\left(1-\sqrt{\gamma}\right)^2\right\} \text{ eventually}\right) = 1.$$

On the other hand, if $\kappa_{\max} \ge \kappa\left(1-\sqrt{\gamma}\right)^{-2}$, then

$$\left\{l_1 \le \lambda_1\left(1+\sqrt{\gamma}\right)^2,\ l_p \ge \lambda_p\left(1-\sqrt{\gamma}\right)^2\right\} \subset \left\{\max\left(\frac{l_1}{\lambda_p}, \frac{\lambda_1}{l_p}\right) \le \kappa_{\max}\right\}.$$

Also, if $\max\left(l_1/\lambda_p,\ \lambda_1/l_p\right) \le \kappa_{\max}$, then

$$\lambda_1/\kappa_{\max} \le l_p \le l_1/\kappa_{\max} \le \lambda_p.$$

From (12), $\tau^\star$ lies between l_p and $l_1/\kappa_{\max}$. Therefore,

$$\tau^\star \le \lambda_p \quad \text{and} \quad \lambda_1 \le \kappa_{\max}\tau^\star.$$

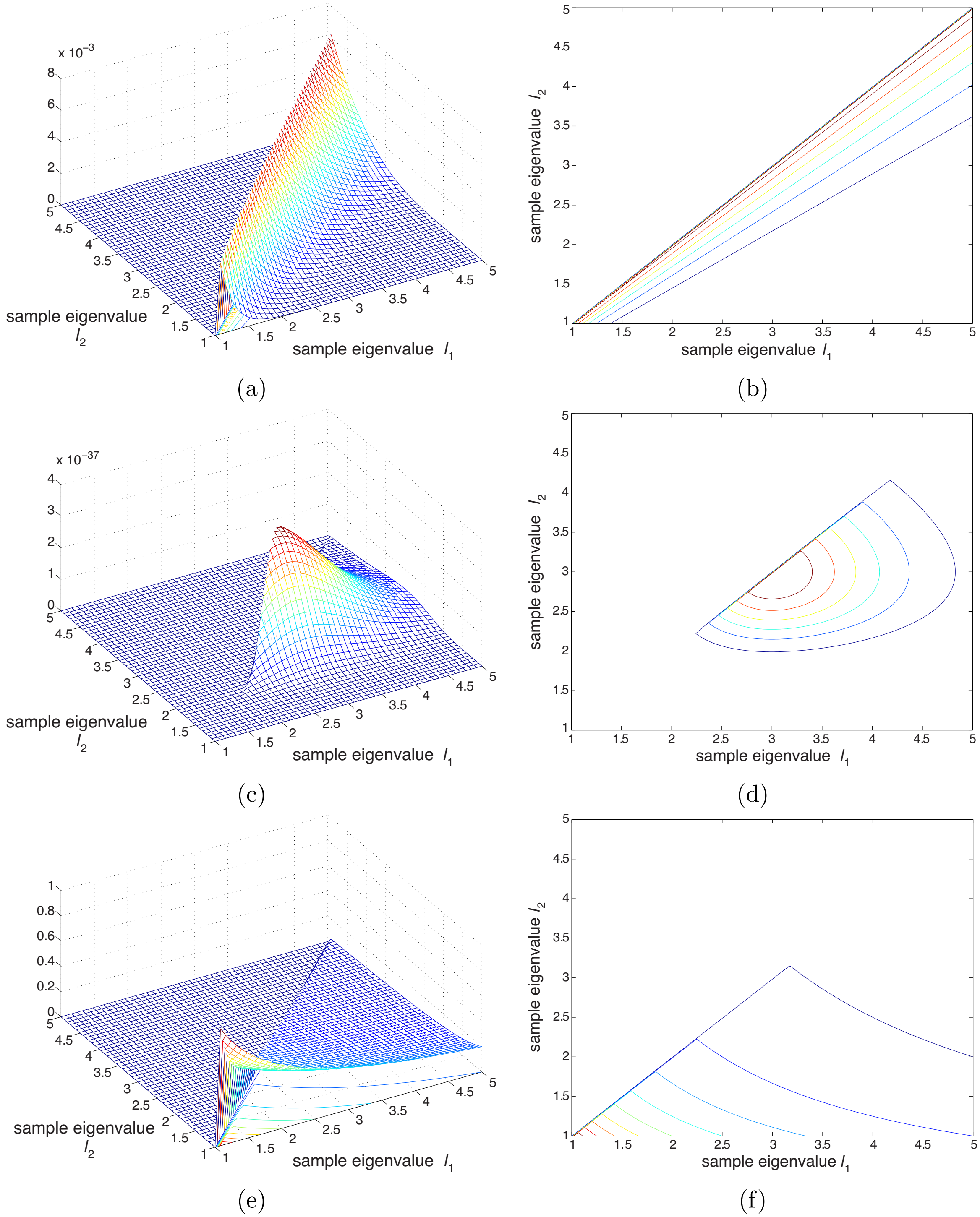
In other words

$$\left\{\max\left(\frac{l_1}{\lambda_p}, \frac{\lambda_1}{l_p}\right) \le \kappa_{\max}\right\} \subset \left\{\Sigma \in \mathcal{D}\left(\kappa_{\max}, u^\star\right)\right\},$$

which concludes the proof.

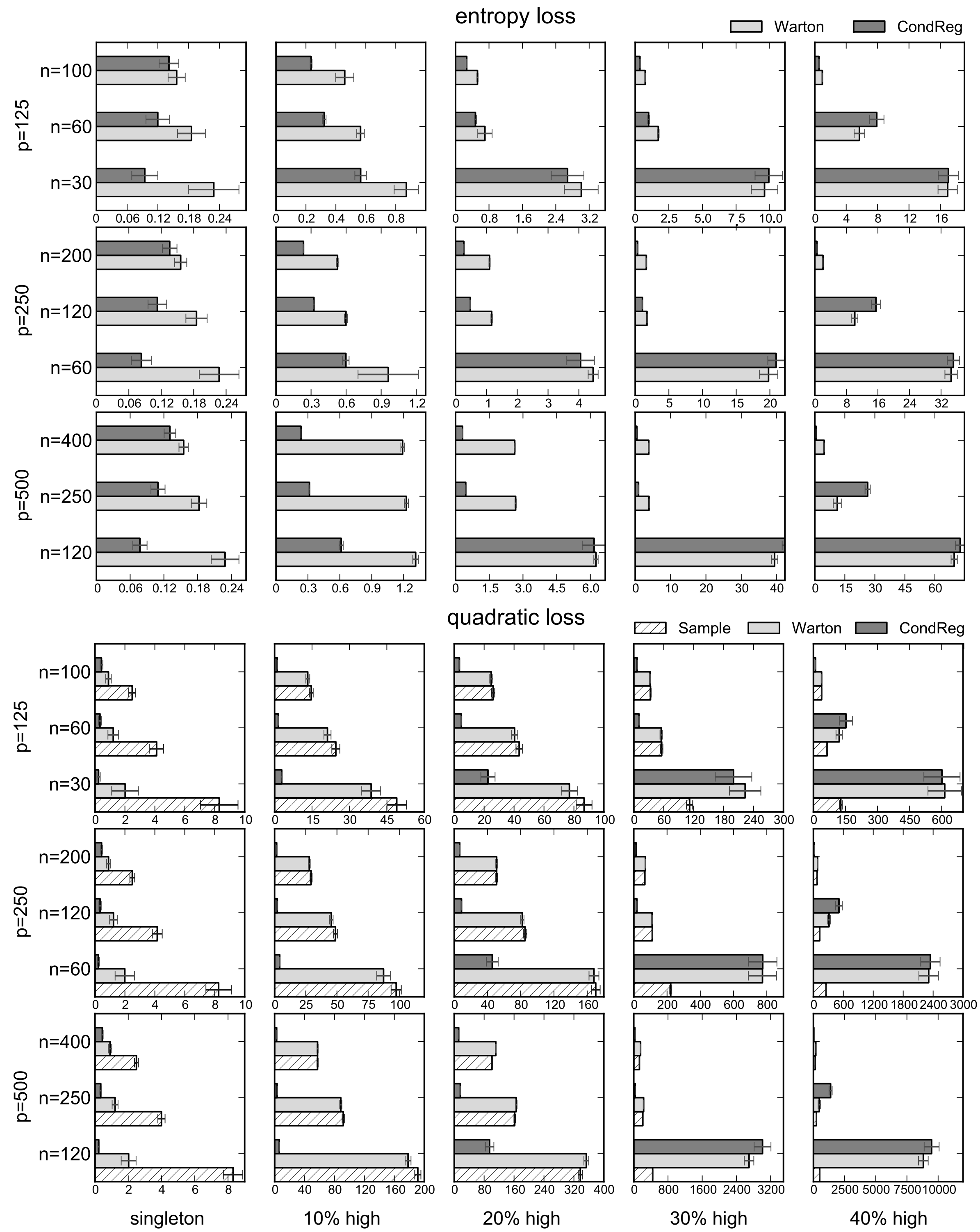
□

B Comparison of Bayesian prior densities



Comparison of various prior densities for eigenvalue shrinkage ($p = 2$). (a), (c), (e): Three-dimensional views. (b), (d), (f): contour views. (a), (b): prior for regularization (19). (c), (d): inverse Wishart prior. (e), (f): reference prior due to Yang and Berger (1994).

C Risk simulation results ($\rho = 0.5$)



Average risks (with error bars) over 1000 runs with respect to two loss functions $\rho = 0.5$. `sample`=sample covariance matrix, `Warton`=linear shrinkage (Warton, 2008), `CondReg`=condition number regularization. Risks are normalized by the dimension (p).

D Empirical minimum variance rebalancing study (Section 6.2)

D.1 List of the Dow Jones constituents

The Dow Jones stocks used in our numerical study and their market performance over the period from February 18, 1994 to June 6, 2008. The return, risk and the Sharpe ratio (SR) are annualized.

index	company	ticker	return [%]	risk [%]	SR
1	3M Company	MMM	12.04	10.74	0.25
2	Alcoa, Inc.	AA	16.50	15.47	0.30
3	American Express	AXP	17.52	14.61	0.35
4	American International Group, Inc.	AIG	7.96	12.93	0.07
5	AT&T Inc.	T	11.57	12.95	0.19
6	Bank of America	BAC	11.57	13.14	0.19
7	The Boeing Company	BA	13.03	13.81	0.23
8	Caterpillar Inc.	CAT	18.53	14.26	0.39
9	Chevron Corporation	CVX	15.86	10.53	0.42
10	Citigroup Inc.	C	14.44	15.27	0.25
11	The Coca-Cola Company	KO	10.74	10.77	0.20
12	E.I. du Pont de Nemours & Company	DD	9.58	12.43	0.13
13	Exxon Mobil Corporation	XOM	16.58	10.46	0.45
14	General Electric Company	GE	13.47	12.04	0.28
15	General Motors Corporation	GM	-1.24	15.85	-0.20
16	The Hewlett-Packard Company	HPQ	20.22	18.24	0.35
17	The Home Depot	HD	12.96	15.28	0.20
18	Intel Corporation	INTC	20.84	19.13	0.35
19	International Business Machines Corp.	IBM	20.99	13.86	0.48
20	Johnson & Johnson	JNJ	17.13	10.10	0.49
21	JPMorgan Chase & Co.	JPM	15.84	15.44	0.29
22	McDonald's Corporation	MCD	14.05	12.05	0.30
23	Merck & Co., Inc.	MRK	12.86	12.87	0.24
24	Microsoft Corporation	MSFT	22.91	15.13	0.50
25	Pfizer Inc.	PFE	15.34	12.92	0.32
26	The Procter & Gamble Company	PG	15.25	11.06	0.37
27	United Technologies Corporation	UTX	18.93	12.37	0.47
28	Verizon Communications Inc.	VZ	9.93	12.38	0.14
29	Wal-Mart Stores, Inc.	WMT	14.86	13.16	0.30
30	The Walt Disney Company	DIS	10.08	14.05	0.13

D.2 Trading periods

index	period	index	period
1	12/14/1992 – 3/31/1993	2	4/1/1993 – 7/19/1993
3	7/20/1993 – 11/2/1993	4	11/3/1993 – 2/17/1994
5	2/18/1994 – 6/8/1994	6	6/9/1994 – 9/23/1994
7	9/26/1994 – 1/11/1995	8	1/12/1995 – 4/28/1995
9	5/1/1995 – 8/15/1995	10	8/16/1995 – 11/30/1995
11	12/1/1995 – 3/19/1996	12	3/20/1996 – 7/5/1996
13	7/8/1996 – 10/21/1996	14	10/22/1996 – 2/6/1997
15	2/7/1997 – 5/27/1997	16	5/28/1997 – 9/11/1997
17	9/12/1997 – 2/29/1997	18	12/30/1997 – 4/17/1998
19	4/20/1998 – 8/4/1998	20	8/5/1998 – 11/18/1998
21	11/19/1998 – 3/10/1999	22	3/11/1999 – 6/25/1999
23	6/28/1999 – 10/12/1999	24	10/13/1999 – 1/28/2000
25	1/31/2000 – 5/16/2000	26	5/17/2000 – 8/31/2000
27	9/1/2000 – 12/18/2000	28	12/19/2000 – 4/6/2001
29	4/9/2001 – 7/25/2001	30	7/26/2001 – 11/14/2001
31	11/15/2001 – 3/6/2002	32	3/7/2002 – 6/21/2002
33	6/24/2002 – 10/8/2002	34	10/9/2002 – 1/27/2003
35	1/28/2003 – 5/14/2003	36	5/15/2003 – 8/29/2003
37	9/2/2003 – 12/16/2003	38	12/17/2003 – 4/5/2004
39	4/6/2004 – 7/23/2004	40	7/26/2004 – 11/8/2004
41	11/9/2004 – 2/25/2005	42	2/28/2005 – 6/14/2005
43	6/15/2005 – 9/29/2005	44	9/30/2005 – 1/18/2006
45	1/19/2006 – 5/5/2006	46	5/8/2006 – 8/22/2006
47	8/23/2006 – 12/7/2006	48	12/8/2006 – 3/29/2007
49	3/30/2007 – 7/17/2007	50	7/18/2007 – 10/31/2007
51	11/1/2007 – 2/20/2008	52	2/21/2008 – 6/6/2008

D.3 Performance metrics

We use the following quantities in assessing the performance of the MVR strategies.

- *Realized return.* The realized return of a portfolio rebalancing strategy over the entire trading period is computed as

$$r_p = \frac{1}{K}\sum_{j=1}^{K}\frac{1}{L}\sum_{t=N_{\text{estim}}+1+(j-1)L}^{N_{\text{estim}}+jL} r^{(t)T}w^{(j)}.$$

- *Realized risk.* The realized risk (return standard deviation) of a portfolio rebalancing strategy over the entire trading period is computed as

$$\sigma_p = \sqrt{\frac{1}{K}\sum_{j=1}^{K}\frac{1}{L}\sum_{t=N_{\text{estim}}+1+(j-1)L}^{N_{\text{estim}}+jL} (r^{(t)T}w^{(j)})^2 - r_p^2}.$$

- *Realized Sharpe ratio (SR).* The realized Sharpe ratio, *i.e.*, the ratio of the excess expected return of a portfolio rebalancing strategy relative to the risk-free return r_f is given by

$$SR = \frac{r_p - r_f}{\sigma_p}.$$

- *Turnover.* The turnover from the portfolio $w^{(j)}$ held at the start date of the jth holding period $[N_{\text{estim}} + 1 + (j-1)L, N_{\text{estim}} + jL]$ to the portfolio $w^{(j-1)}$ held at the previous period is computed as

$$\text{TO}(j) = \sum_{i=1}^{p} \left| w_i^{(j)} - \left(\prod_{t=N_{\text{estim}}+1+(j-1)L}^{N_{\text{estim}}+jL} r_i^{(t)} \right) w_i^{(j-1)} \right|.$$

  For the first period, we take $w^{(0)} = 0$, *i.e.*, the initial holdings of the assets are zero.

- *Normalized wealth growth.* Let $w^{(j)} = (w_1^{(j)}, \ldots, w_n^{(j)})$ be the portfolio constructed by a rebalancing strategy held over the jth holding period $[N_{\text{estim}} + 1 + (j-1)L, N_{\text{estim}} + jL]$. When the initial budget is normalized to one, the normalized wealth grows according to the recursion

$$W(t) = \begin{cases} W(t-1)(1 + \sum_{i=1}^{p} w_{it} r_i^{(t)}), & t \notin \{N_{\text{estim}} + jL \mid j = 1, \ldots, K\}, \\ W(t-1)(1 + \sum_{i=1}^{p} w_{it} r_i^{(t)}) - \text{TC}(j), & t = N_{\text{estim}} + jL, \end{cases}$$

  for $t = N_{\text{estim}}, \ldots, N_{\text{estim}} + KL$, with the initial wealth $W(N_{\text{estim}}) = 1$. Here

$$w_{it} = \begin{cases} w_i^{(1)}, & t = N_{\text{estim}} + 1, \ldots, N_{\text{estim}} + L, \\ \vdots & \\ w_i^{(K)}, & t = N_{\text{estim}} + 1 + (K-1)L, \ldots, N_{\text{estim}} + KL. \end{cases}$$

  and

$$\text{TC}(j) = \sum_{i=1}^{p} \eta_i \left| w_i^{(j)} - \left(\prod_{t=N_{\text{estim}}+1+(j-1)L}^{N_{\text{estim}}+jL} r_i^{(t)} \right) w_i^{(j-1)} \right|$$

  is the transaction cost due to the rebalancing if the cost to buy or sell one share of stock i is η_i.

- *Size of the short side.* The size of the short side of a portfolio rebalancing strategy over the entire trading period is computed as

$$r_p = \frac{1}{K} \sum_{j=1}^{K} \left(\sum_{i=1}^{p} |\min(w_i^{(j)}, 0)| / \sum_{i=1}^{p} |w_i^{(j)}| \right).$$

D.4 Minimum variance portfolio theoretic performance metrics

Realized utility, realized return, and realized risk based on different regularization schemes for covariance matrices are reported. `sample`=sample covariance matrix, `LW`=linear shrinkage (Ledoit and

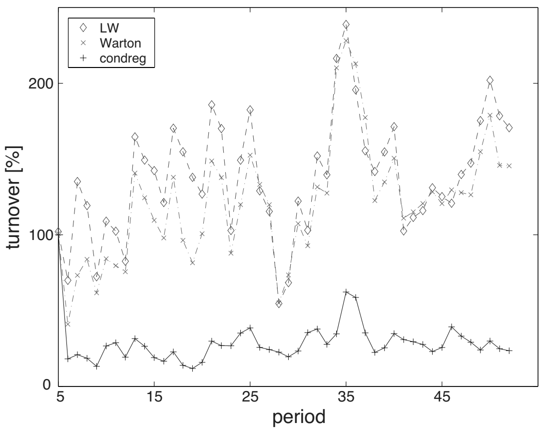
Wolf, 2004), `Warton`=linear shrinkage (Warton, 2008), `condreg`=condition number regularization.

Each entry is the mean (standard deviation) of the corresponding metric over the trading period (48 holding periods) from February 1994 through June 2008. The standard deviations are computed in a heteroskedasticity-and-autocorrelation consistent manner discussed by Ledoit and Wolf (2008, Sec. 3.1.). For comparison, the performance metrics of the S&P 500 index for the same period are reported at the end of the table. All values are annualized.

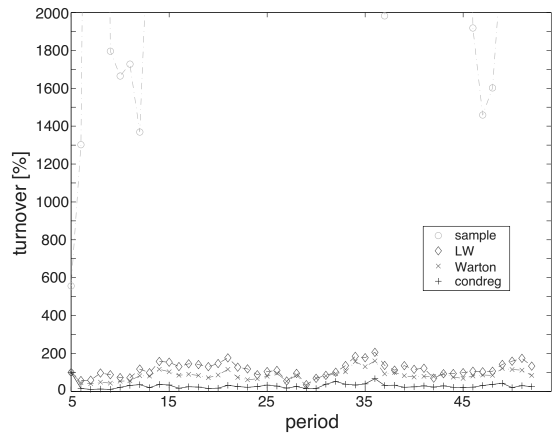
covariance regularization scheme	return [%]	risk [%]	SR
		$N_{\text{estim}} = 15$	
`sample`	–	–	–
`LW`	12.12 (3.69)	14.38 (0.69)	0.50 (0.26)
`Warton`	12.21 (3.71)	14.28 (0.68)	0.50 (0.26)
`condreg`	14.49 (4.01)	16.07 (0.94)	0.59 (0.25)
`condreg - LW`	2.37 (5.45)	1.69 (1.17)	0.10 (0.36)
`condreg - Warton`	2.28 (5.47)	1.79 (1.16)	0.09 (0.37)
		$N_{\text{estim}} = 30$	
`sample`	145.91 (114.74)	851.07 (257.28)	0.17 (0.14)
`LW`	11.82 (3.72)	14.27 (0.65)	0.48 (0.26)
`Warton`	12.1 (3.68)	14.06 (0.69)	0.51 (0.27)
`condreg`	14.32 (4.00)	15.92 (0.91)	0.59 (0.26)
`condreg - LW`	2.51 (5.46)	1.65 (1.12)	0.11 (0.37)
`condreg - Warton`	2.16 (5.43)	1.85 (1.14)	0.08 (0.37)
		$N_{\text{estim}} = 45$	
`sample`	9.52 (5.62)	21.15 (0.74)	0.21 (0.26)
`LW`	12.21 (3.65)	14.04 (0.72)	0.51 (0.26)
`Warton`	12.62 (3.60)	13.85 (0.76)	0.55 (0.26)
`condreg`	14.53 (3.93)	15.63 (0.87)	0.61 (0.26)
`condreg - LW`	2.33 (5.36)	1.60 (1.13)	0.10 (0.37)
`condreg - Warton`	1.91 (5.33)	1.79 (1.16)	0.06 (0.37)
		$N_{\text{estim}} = 60$	
`sample`	9.32 (4.29)	17.52 (0.70)	0.25 (0.25)
`LW`	11.55 (3.55)	13.94 (0.74)	0.47 (0.26)
`Warton`	12.15 (3.55)	13.78 (0.77)	0.52 (0.26)
`condreg`	14.13 (3.91)	15.65 (0.90)	0.58 (0.25)
`condreg - LW`	2.58 (5.28)	1.71 (1.17)	0.11 (0.36)
`condreg - Warton`	1.98 (5.28)	1.87 (1.18)	0.06 (0.37)
		S&P500	
–	8.71 (3.84)	16.04 (0.94)	0.23 (0.24)

D.5 Turnover

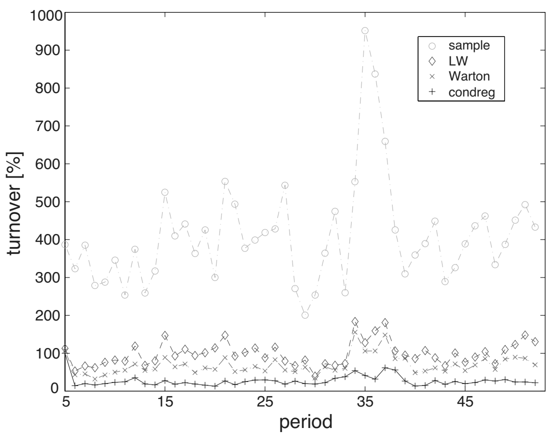
Turnover for various estimation horizon sizes over the trading period from February 18, 1994 through June 6, 2008. `sample`=sample covariance matrix, `LW`=linear shrinkage (Ledoit and Wolf, 2004), `Warton`=linear shrinkage (Warton, 2008), `condreg`=condition number regularization.



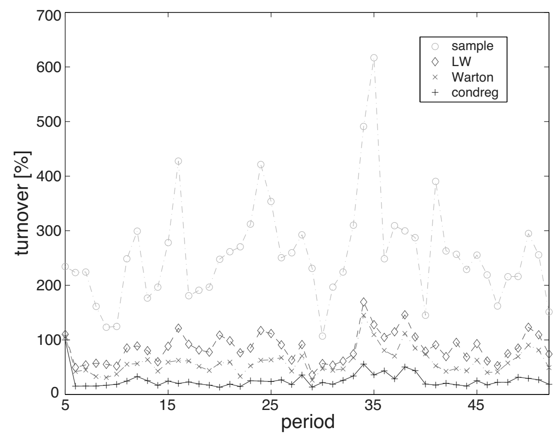
(a) $N_{\text{estim}} = 15$



(b) $N_{\text{estim}} = 30$



(c) $N_{\text{estim}} = 45$



(d) $N_{\text{estim}} = 60$

D.6 Size of the short side

The size of the short side of the portfolios based on different regularization schemes for covariance matrices are reported. `sample`=sample covariance matrix, `LW`=linear shrinkage (Ledoit and Wolf, 2004), `Warton`=linear shrinkage (Warton, 2008), `condreg`=condition number regularization.

Each entry is the mean (standard deviation) of the short-side size over the trading period (48 holding periods) from February 1994 through June 2008, given in percent.

covariance	N_{estim}			
regularization scheme	15	30	45	60
`sample`	–	47.47 (1.86)	36.84 (3.21)	32.83 (4.47)
`LW`	13.05 (7.68)	15.87 (7.63)	17.29 (7.35)	17.79 (7.36)
`Warton`	10.68 (7.62)	11.01 (6.91)	11.14 (7.09)	11.53 (0.73)
`condreg`	0.08 (0.54)	0.27 (0.94)	0.61 (1.88)	0.96 (2.81)

References

Geman, S. (1980). A limit theorem for the norm of random matrices. *The Annals of Probability 8*(2), 252–261.

Ledoit, O. and M. Wolf (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis 88*, 365–411.

Ledoit, O. and M. Wolf (2008, December). Robust performance hypothesis testing with the sharpe ratio. *Journal of Empirical Finance 15*(5), 850–859.

Silverstein, J. W. (1985). The smallest eigenvalue of a large dimensional wishart matrix. *The Annals of Probability 13*(4), 1364–1368.

Warton, D. I. (2008). Penalized Normal Likelihood and Ridge Regularization of Correlation and Covariance Matrices. *Journal of the American Statistical Association 103*(481), 340–349.

Yang, R. and J. O. Berger (1994). Estimation of a covariance matrix using the reference prior. *The Annals of Statistics 22*(3), 1195–1211.