Details for the discussion of “Large covariance estimation by thresholding principal orthogonal complements”

Hajo Holzmann[1] **and Anna Leister**

Fachbereich Mathematik und Informatik, Philipps-Universität Marburg, Germany.

Abstract

We provide the details for the results stated in our contribution to the discussion of “Large covariance estimation by thresholding principal orthogonal complements” by Fan, Liao and Mincheva (2013). In particular, we present the rates for estimating the covariance matrix as well as the lagged covariance matrix in high-dimensional, approximate factor models when using a. the empirical covariance matrix, b. an estimate based on the observed factors, c. an estimate based on estimated factors. Further, we provide some simulations results and describe the simulation setting in detail.

The recent discussion paper by Fan, Liao and Mincheva (2013) (FLM in the following) on “ Large covariance estimation by thresholding principal orthogonal complements” in Series B provides important results on estimating the factor structure in high-dimensional, approximate factor models, and its implications for estimating the underlying covariance matrix. Here we consider implications for the time series structure of the observed series $(\boldsymbol{y}_t)$, specifically its lagged covariance matrix, and convergence in the $\|\cdot\|_{\mathrm{MAX}}$-norm, defined by $\|A\|_{\mathrm{MAX}} = \max_{i,j} |a_{i,j}|$.

Let us briefly recall the notation and some assumptions from FLM. Consider a K-factor model

$$\boldsymbol{y}_t = \boldsymbol{B}\boldsymbol{f}_t + \boldsymbol{u}_t, \tag{1}$$

where $\boldsymbol{y}_t$ is a vector of p observations at time $t \in \{1,\dots,T\}$, $\boldsymbol{B} = (\boldsymbol{b}_1,\dots,\boldsymbol{b}_p)^T$ is a $p\times K$ matrix of factor loadings ($\boldsymbol{b}_i \in \mathbb{R}^K$, $i \in \{1,\dots,p\}$), $\boldsymbol{f}_t$ is a $K\times 1$ vector of common factors at time t and $\boldsymbol{u}_t$ is a $p\times 1$ vector of error terms at time t. For simplicity, we assume the means to be removed ($E(y_{it}) = E(f_{jt}) = 0$).
If $\boldsymbol{f}_t$ and $\boldsymbol{u}_t$ are uncorrelated, we have

$$\Sigma := \mathrm{cov}(\boldsymbol{y}_t) = \boldsymbol{B}\mathrm{cov}(\boldsymbol{f}_t)\boldsymbol{B}' + \mathrm{cov}(\boldsymbol{u}_t). \tag{2}$$

The following assumptions are imposed in (see Fan et al., 2011a,b).

Assumption 1. $\{\boldsymbol{f}_t, \boldsymbol{u}_t\}_{t\geq 1}$ is a strictly stationary process with $E(f_{jt}, u_{it}) = 0$ for all $i = 1,\dots,p$, $j = 1,\dots,K$ and $t = 1,\dots,T$.

[1]Address for correspondence: Prof. Dr. Hajo Holzmann, Philipps-Universität Marburg, Fachbereich Mathematik und Informatik, Hans-Meerwein-Straße. D-35032 Marburg, Germany, Email: holzmann@mathematik.uni-marburg.de, Fon: + 49 6421 2825454

Assumption 2. For $\{\boldsymbol{u}_t\}_{t\geq 1}$ we have $E(u_{it}) = 0$ for all $i = 1,\dots,p$. There exists a constant $c_1 > 0$ such that for the minimum eigenvalue of the covariance matrix we have $c_1 < \lambda_{\min}(\mathrm{cov}(\boldsymbol{u}_t))$.
In addition, there are constants r_1, $b_1 > 0$ such that for all $s > 0$

$$P(|u_{it}| > s) \leq \exp\big(-(\frac{s}{b_1})^{r_1}\big), \qquad i = 1,\dots,p.$$

Assumption 3. There exist r_2, $b_2 > 0$ such that for all $s > 0$

$$P(|f_{jt}| > s) \leq \exp\big(-(\frac{s}{b_2})^{r_2}\big), \qquad j = 1,\dots,K.$$

We denote by $\alpha(T)$ the mixing coefficient

$$\alpha(T) = \sup_{A\in\mathcal{F}^0_{-\infty},\ B\in\mathcal{F}^\infty_T} |P(A)P(B) - P(AB)|,$$

where $\mathcal{F}^0_{-\infty} = \sigma\{(\boldsymbol{f}_t, \boldsymbol{u}_t):\ t \leq 0\}$ and $\mathcal{F}^\infty_T = \sigma\{(\boldsymbol{f}_t, \boldsymbol{u}_t):\ t \geq T\}$.

Assumption 4. There exist $C, r_3 > 0$ with $3r_1^{-1} + 1.5r_2^{-1} + r_3^{-1} > 0$ such that

$$\alpha(T) \leq \exp(-CT^{r_3})$$

for all $T \in \mathbb{Z}^+$.

Assumption 5. As $p \to \infty$ we have

$$0 < \lambda_{\min}(p^{-1}\boldsymbol{B}^T\boldsymbol{B}) \leq \lambda_{\max}(p^{-1}\boldsymbol{B}^T\boldsymbol{B}) < \infty.$$

Assumption 6. We have $\max_{1\leq i\leq p}\|\boldsymbol{b}_i\| = O(1)$. In addition, there is a constant $M > 0$ satisfying

$$E\big(p^{-\frac{1}{2}}(\boldsymbol{u}_s'\boldsymbol{u}_t - E(\boldsymbol{u}_s'\boldsymbol{u}_t))\big)^4 < M \qquad \text{and}$$
$$E\|p^{-\frac{1}{2}}\sum_{i=1}^p \boldsymbol{b}_i u_{it}\|^4 < M.$$

Assumption 7. The number of common factors K is fixed whereas p and T diverge to ∞ satisfying $T = o(p^2)$ and $\log(p) = o(T^{\gamma/6})$ where $\gamma^{-1} = 3r_1^{-1} + 1.5r_2^{-1} + r_3^{-1} + 1$.

Estimating the covariance matrix

First let us consider estimation of Σ in (2).
¿From Assumptions 1 - 4, FLM obtain the following results

$$\max_{1\leq i,j\leq K}|\frac{1}{T}\sum_{t=1}^T f_{it}f_{jt} - E(f_{i1}f_{j1})| = O_P(T^{-\frac{1}{2}}) \tag{3}$$

$$\max_{1\leq i,j\leq p}|\frac{1}{T}\sum_{t=1}^T u_{it}u_{jt} - E(u_{i1}u_{j1})| = O_P(\sqrt{\log p/T}) \tag{4}$$

$$\max_{1\leq i\leq K, 1\leq j\leq p}|\frac{1}{T}\sum_{t=1}^T f_{it}u_{jt}| = O_P(\sqrt{\log p/T}). \tag{5}$$

Now, utilizing a *known* factor structure, Fan et al. (2011a, Theorem 3.2) obtain the rate $O_P\big(\sqrt{(\log p)/T}\big)$ for estimating Σ in $\|\cdot\|_{\mathrm{MAX}}$. Further, in case of an unobserved factor structure, FLM obtain for an estimate of Σ based on estimated factors the rate $O_P\big(1/\sqrt{p}+\sqrt{(\log p)/T}\big)$.
Consider now the empirical covariance matrix. Writing

$$\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{y}_t\boldsymbol{y}_t' - \Sigma = \boldsymbol{B}\frac{1}{T}\sum_{t=1}^{T}\big(\boldsymbol{f}_t\boldsymbol{f}_t' - \mathrm{cov}(\boldsymbol{f}_t)\big)\boldsymbol{B}' + \frac{1}{T}\sum_{t=1}^{T}\big(\boldsymbol{u}_t\boldsymbol{u}_t' - \mathrm{cov}(\boldsymbol{u}_t)\big)$$
$$+\boldsymbol{B}\frac{1}{T}\sum_{t=1}^{T}\big(\boldsymbol{f}_t\boldsymbol{u}_t'\big) + \frac{1}{T}\sum_{t=1}^{T}\big(\boldsymbol{u}_t\boldsymbol{f}_t'\big)\boldsymbol{B}'$$

Use (3) - (5) and assumption 6 to obtain the rate $O_P\big(\sqrt{(\log p)/T}\big)$.
Thus, there is no theoretical gain in using the factor structure from these rates. For finite sample results, see Figure 3 below.

Estimating the lagged covariance

Next, we investigate estimation of $\mathrm{cov}(\boldsymbol{y}_t,\boldsymbol{y}_{t+h}) = E\boldsymbol{y}_t\boldsymbol{y}_{t+h}'$.
For distinction, assume that the errors $(\boldsymbol{u}_t)$ are known to be serially uncorrelated:

$$\mathrm{cov}(\boldsymbol{y}_t,\boldsymbol{y}_{t+h}) = \boldsymbol{0}, \qquad h\neq 0.$$

In extension of (3) - (5), we have that for $0<\alpha<1$, under the above assumptions

$$\max_{1\le h\le T^\alpha}\max_{1\le i,j\le K}\Big|\frac{1}{T}\sum_{t=1}^{T-h} f_{it}\, f_{j(t+h)} - E(f_{i1}f_{j(1+h)})\Big| = O_P(T^{-\frac{1}{2}})$$
$$\max_{1\le h\le T^\alpha}\max_{1\le i,j\le p}\Big|\frac{1}{T}\sum_{t=1}^{T-h} u_{it}\, u_{jt} - E(u_{i1}u_{j1})\Big| = O_P(\sqrt{\log p/T})$$
$$\max_{1\le h\le T^\alpha}\max_{1\le i\le K,1\le j\le p}\Big|\frac{1}{T}\sum_{t=1}^{T-h} f_{it}\, u_{j(t+h)}\Big| = O_P(\sqrt{\log p/T}).$$
$$\max_{1\le h\le T^\alpha}\max_{1\le i\le K,1\le j\le p}\Big|\frac{1}{T}\sum_{t=1}^{T-h} f_{i(t+h)}\, u_{j}\Big| = O_P(\sqrt{\log p/T}).$$

Now, for the sample auto-covariance matrix, we have that

$$\frac{1}{T}\sum_{t=1}^{T-h}\boldsymbol{y}_t\boldsymbol{y}_{t+h}' - \mathrm{cov}(\boldsymbol{y}_1,\boldsymbol{y}_{1+h}) = \boldsymbol{B}\frac{1}{T}\sum_{t=1}^{T-h}\big(\boldsymbol{f}_t\boldsymbol{f}_{t+h}' - \mathrm{cov}(\boldsymbol{f}_t,\boldsymbol{f}_{t+h})\big)\boldsymbol{B}' + \frac{1}{T}\sum_{t=1}^{T-h}\big(\boldsymbol{u}_t\boldsymbol{u}_{t+h}'\big)$$
$$+\boldsymbol{B}\frac{1}{T}\sum_{t=1}^{T-h}\big(\boldsymbol{f}_t\boldsymbol{u}_{t+h}'\big) + \frac{1}{T}\sum_{t=1}^{T-h}\big(\boldsymbol{u}_t\boldsymbol{f}_{t+h}'\big)\boldsymbol{B}'$$
$$-\frac{h}{T}\boldsymbol{B}\mathrm{cov}(\boldsymbol{f}_t,\boldsymbol{f}_{t+h})\boldsymbol{B}'.$$

and therefore we obtain the rate $O_P\big(\sqrt{(\log p)/T} + \frac{h}{T}\left\|\mathrm{cov}(\boldsymbol{f}_t, \boldsymbol{f}_{t+h})\right\|_{\mathrm{MAX}}\big)$.
With known factor structure, we have that

$$
\begin{aligned}
&\hat{\boldsymbol{B}}\frac{1}{T}\sum_{t=1}^{T-h}\big(\boldsymbol{f}_t\boldsymbol{f}'_{t+h}\big)\hat{\boldsymbol{B}}' - \boldsymbol{B}\mathrm{cov}(\boldsymbol{f}_1, \boldsymbol{f}_{1+h})\boldsymbol{B}' \\
=&\hat{\boldsymbol{B}}\frac{1}{T}\sum_{t=1}^{T-h}\big(\boldsymbol{f}_t\boldsymbol{f}'_{t+h} - \mathrm{cov}(\boldsymbol{f}_t, \boldsymbol{f}_{t+h})\big)\hat{\boldsymbol{B}}' + \big(\hat{\boldsymbol{B}} - \boldsymbol{B}\big)\mathrm{cov}(\boldsymbol{f}_t, \boldsymbol{f}_{t+h})\big)\hat{\boldsymbol{B}}' \\
&+ \boldsymbol{B}\,\mathrm{cov}(\boldsymbol{f}_t, \boldsymbol{f}_{t+h})\big)\big(\hat{\boldsymbol{B}}' - \boldsymbol{B}'\big) - \frac{h}{T}\boldsymbol{B}\mathrm{cov}(\boldsymbol{f}_t, \boldsymbol{f}_{t+h})\boldsymbol{B}',
\end{aligned}
$$

where $\hat{B}$ is a least-squares estimate of B. Now, Fan, Liao and Mincheva (2011) show that

$$
\max_{1\le i\le K}\|\hat{\boldsymbol{b}}_i - \boldsymbol{b}_i\| = O_P\big(\sqrt{(\log p)/T}\big),
$$

and therefore in summary we again obtain the rate $O_P\big(\sqrt{1/T} + \frac{h}{T}\left\|\mathrm{cov}(\boldsymbol{f}_t, \boldsymbol{f}_{t+h})\right\|_{\mathrm{MAX}}\big)$.

Finally using a factor structure with unknown factors, we employ the estimate given at entry $1 \le i, j \le p$ by

$$
\frac{1}{T}\sum_{t=1}^{T-h}\big(\hat{\boldsymbol{b}}'_i\hat{\boldsymbol{f}}_t\hat{\boldsymbol{f}}'_{t+h}\hat{\boldsymbol{b}}_j\big),
$$

where $\hat{\boldsymbol{b}}'_i\hat{\boldsymbol{f}}_t$ is as in FLM. Indeed, difference

$$
\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T-h}\big(\hat{\boldsymbol{b}}'_i\hat{\boldsymbol{f}}_t\hat{\boldsymbol{f}}'_{t+h}\hat{\boldsymbol{b}}_j\big) - \boldsymbol{b}'_i\mathrm{cov}(\boldsymbol{f}_1, \boldsymbol{f}_{1+h})\boldsymbol{b}_j \\
=&\frac{1}{T}\sum_{t=1}^{T-h}\big(\hat{\boldsymbol{b}}'_i\hat{\boldsymbol{f}}_t\hat{\boldsymbol{f}}'_{t+h}\hat{\boldsymbol{b}}_j - \boldsymbol{b}'_i\boldsymbol{f}_t\boldsymbol{f}'_{t+h}\boldsymbol{b}_j\big) + \frac{1}{T}\boldsymbol{b}'_i\sum_{t=1}^{T-h}\big(\boldsymbol{f}_t\boldsymbol{f}'_{t+h} - \mathrm{cov}(\boldsymbol{f}_1, \boldsymbol{f}_{1+h})\big)\boldsymbol{b}_j \\
&+ h/T\,\boldsymbol{b}'_i\mathrm{cov}(\boldsymbol{f}_1, \boldsymbol{f}_{1+h})\boldsymbol{b}_j.
\end{aligned}
$$

For the first term,

$$
\begin{aligned}
&\Big|\frac{1}{T}\sum_{t=1}^{T-h}\big(\hat{\boldsymbol{b}}'_i\hat{\boldsymbol{f}}_t\hat{\boldsymbol{f}}'_{t+h}\hat{\boldsymbol{b}}_j - \boldsymbol{b}'_i\boldsymbol{f}_t\boldsymbol{f}'_{t+h}\boldsymbol{b}_j\big)\Big| \\
\le&\frac{1}{T}\sum_{t=1}^{T-h}\big|\hat{\boldsymbol{b}}'_i\hat{\boldsymbol{f}}_t - \boldsymbol{b}'_i\boldsymbol{f}_t\big|\|\hat{\boldsymbol{f}}_{t+h}\|\|\hat{\boldsymbol{b}}_j\| + \|\boldsymbol{b}_i\|\|\boldsymbol{f}_t\|\big|\hat{\boldsymbol{f}}'_{t+h}\hat{\boldsymbol{b}}_j - \boldsymbol{f}'_{t+h}\boldsymbol{b}_j\big|.
\end{aligned}
$$

¿From Corollary 1 in FLM and

$$
\max_{1\le j\le p}\big(\|\hat{\boldsymbol{b}}_j\| + \|\boldsymbol{b}_j\|\big) = O_P(1), \qquad \max_{1\le t\le T}\big(\|\hat{\boldsymbol{f}}_t\| + \|\boldsymbol{f}_t\|\big) = O_P((\log T)^{1/r_2})
$$

we get the rate $O_P\big((\log T)^{2/r_2}\sqrt{(\log p)/T} + (\log T)^{1/r_2}T^{1/4}/\sqrt{p}\big)$ for this term, and totally a rate of

$$
O_P\big((\log T)^{2/r_2}\sqrt{(\log p)/T} + (\log T)^{1/r_2}T^{1/4}/\sqrt{p} + \frac{h}{T}\left\|\mathrm{cov}(\boldsymbol{f}_t, \boldsymbol{f}_{t+h})\right\|_{\mathrm{MAX}}\big).
$$

Summary

In terms of the rates of convergence given above, there is no particular gain in using the factor structure. The simulations show that there is some finite-sample gain, in particular for larger values of p.

Further issues that would be of some interest are evulation in distinct norms, estimation of the cross-correlation matrix as well as the long-range autocovariance matrix.

Simulations.

Our simulations are similar to those in Fan et al. (2008, 2011a)). Specifically, we simulate an approximate factor model with $K = 3$ and $T = 500$ fixed and dimension p growing from 20 to 600 with increment 20. For calibration we use the Fama-French three factor model (see Fama and French, 1992). The data of 30 industry portfolios are available from `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html` (15.11.2012) to fit the three factor model for the observed factors and use the estimated factor loadings to estimate mean $\hat{\mu}_b$ and covariance $\hat{\text{cov}}(\boldsymbol{b}_i)$ as well as the standard deviation of the residuals $\hat{\sigma}_i$, $i = 1, \dots, p$ for simulating $\text{cov}(\boldsymbol{u}_t)$.

We then simulate factor loadings from $\mathcal{N}(\hat{\mu}_b, \hat{\text{cov}}(\boldsymbol{b}_i))$, create a sparse positive definite matrix $\text{cov}(\boldsymbol{u}_t)$ using $\hat{\sigma}_i$ and simulate error terms from $\mathcal{N}(\boldsymbol{0}, \text{cov}(\boldsymbol{u}_t))$.

In order to obtain more stringly dependent factors, we simulate from an AR(1) model $\boldsymbol{f}_t = \Phi \boldsymbol{f}_{t-1} + \varepsilon_t$ with

$$\Phi = \begin{pmatrix} 0.75 & 0.1 & -0.03 \\ 0.05 & 0.8 & -0.015 \\ 0.01 & -0.05 & 0.6 \end{pmatrix}.$$

We then compose the observations $\boldsymbol{y}_t$ by using the simulated components of the factor model and obtain the lagged covariances $\Sigma(h) := \text{cov}(\boldsymbol{y}_t, \boldsymbol{y}_{t+h}) = \boldsymbol{B}\text{cov}(\boldsymbol{f}_t, \boldsymbol{f}_{t+h})\boldsymbol{B}'$. For estimation of $\Sigma(h)$ we use three different approaches:

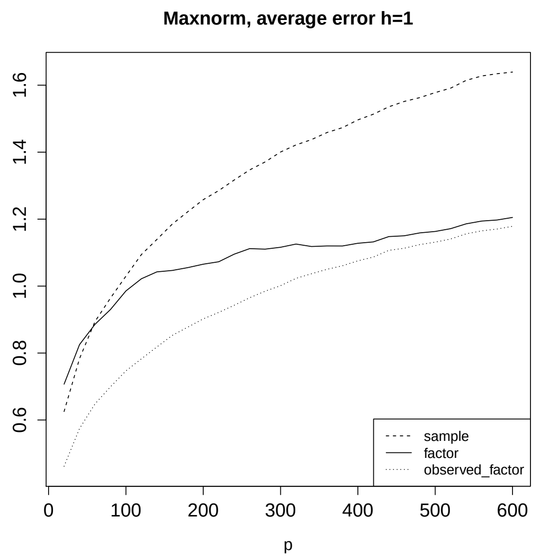
(i) $\hat{\Sigma}_{sam}(h)$: Using the simulated values of $\boldsymbol{y}_t$ we calculate the sample autocovariance according to

$$\begin{aligned} \hat{\Sigma}_{sam}(h) &= \frac{1}{T}\sum_{t=1}^{T-h}(\boldsymbol{y}_t - \bar{\boldsymbol{y}})(\boldsymbol{y}_{t+h} - \bar{\boldsymbol{y}})', \\ \bar{\boldsymbol{y}} &= \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{y}_t. \end{aligned} \tag{6}$$

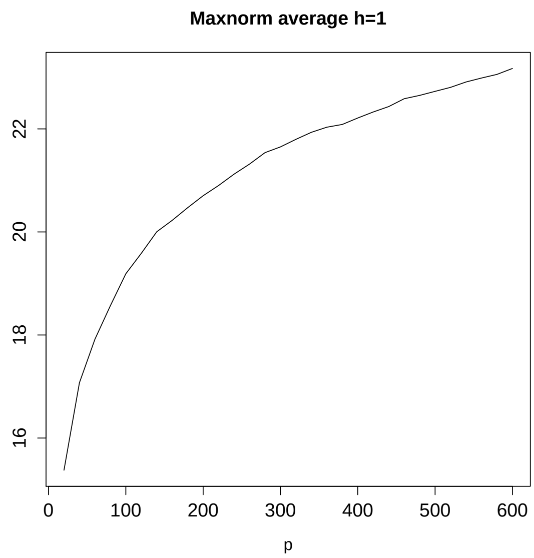
(ii) $\hat{\Sigma}_{obs}(h)$: We use the factor model with observed factors to estimate $\boldsymbol{B}$ by using the least squares method and estimate $\hat{\text{cov}}(\boldsymbol{f}_t, \boldsymbol{f}_{t+h})$ by (6). Therefore we have $\hat{\Sigma}_{obs}(h) = \hat{\boldsymbol{B}}_{LS}\hat{\text{cov}}(\boldsymbol{f}_t, \boldsymbol{f}_{t+h})_{obs}\hat{\boldsymbol{B}}'_{LS}$.

(iii) $\hat{\Sigma}(h)$: To apply the factor model with unobserved factors we use the POET package (see Fan et al., 2011b) to estimate the factor loadings and the factors. We then apply (6) to obtain $\hat{\Sigma}(h) = \hat{\boldsymbol{B}}\hat{\text{cov}}(\boldsymbol{f}_t, \boldsymbol{f}_{t+h})\hat{\boldsymbol{B}}'$.

We repeat the simulation 500 times and plot the averages of the distance from $\hat{\Sigma}_{sam}(h)$, $\hat{\Sigma}_{obs}(h)$ and $\hat{\Sigma}(h)$ to $\Sigma(h)$ under the MAX-norm for $h = 0, 1, 5$, see Figures 1 - 3. There is some finite-sample gain for higher-dimensional p when using the factor structure.

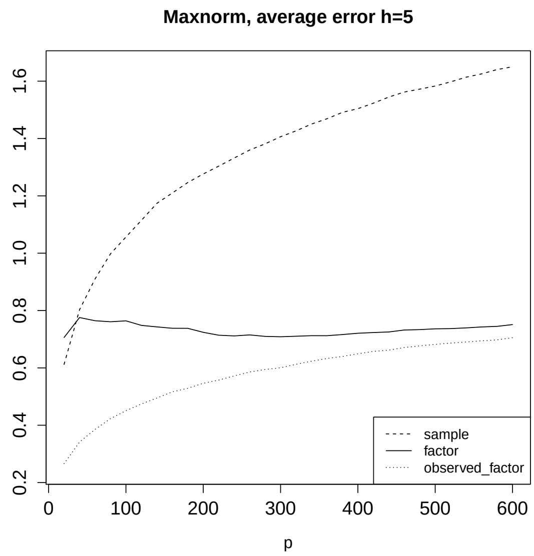


(a)

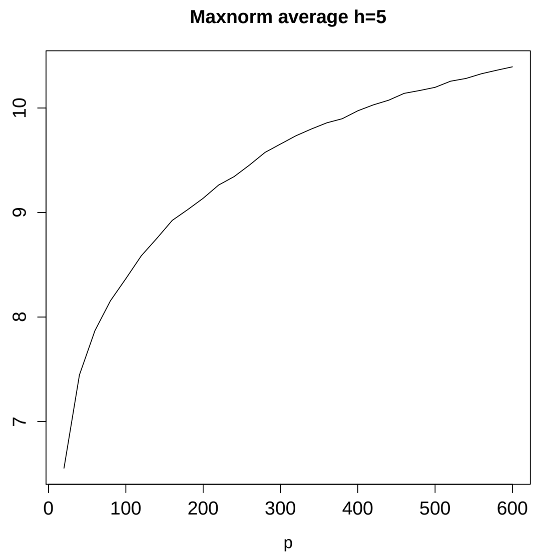


(b)

Figure 1: Timelag h=1: 1(a) Averages of errors (max norm) over 500 simulations against p. 1(b) Averages of lagged covariance (max norm) over 500 simulations against p.

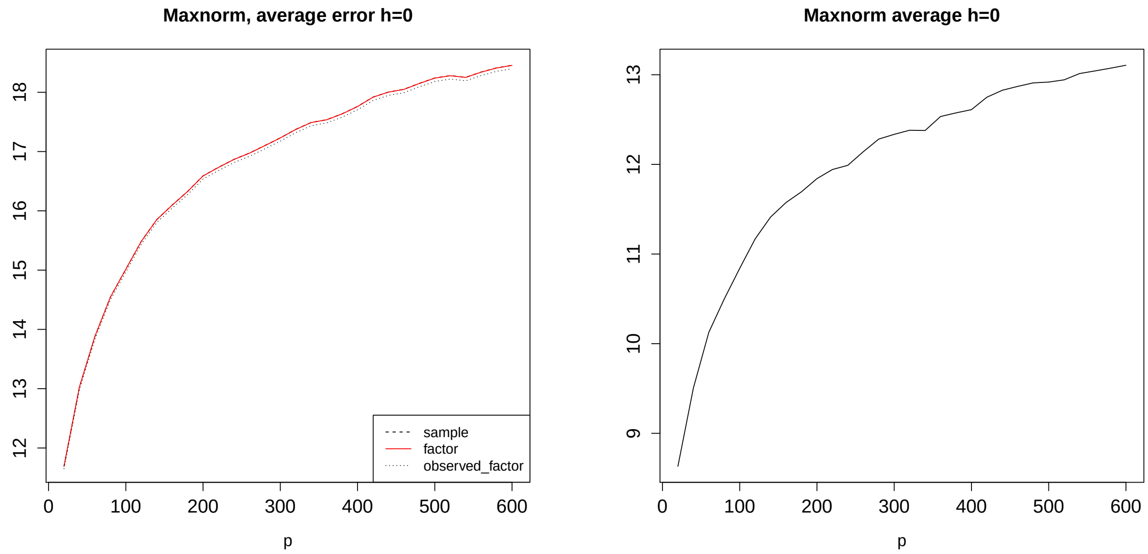


(a)



(b)

Figure 2: Timelag h=5: 2(a) Averages of errors (max norm) over 500 simulations against p. 1(b) Averages of lagged covariance (max norm) over 500 simulations against p.



(a) (b)

Figure 3: Timelag h=0: 3(a) Averages of errors (max norm) over 500 simulations against p. 3(b) Averages of covariance (max norm) over 500 simulations against p.

References

E. F. Fama and K. R. French. The cross-section of expected stock returns. *Journal of Finance*, 47(2):427–65, June 1992.

J. Fan, Y. Fan, and J. Lv. High dimensional covariance matrix estimation using a factor model. *Journal of Econometrics*, 147(1):186–197, November 2008.

J. Fan, Y. Liao, and M. Mincheva. High-dimensional covariance matrix estimation in approximate factor models. *Ann. Statist.*, 39:3320–3356, 2011a.

J. Fan, Y. Liao, and M. Mincheva. Large covariance estimation by thresholding principal orthogonal complements. *Available at SSRN 1977673*, 2011b.