Summary of factor model analysis by Fan at al.

W. Evan Durno

July 11, 2015

Brief summary

Factor models for high-dimensional data are analysed [Fan et al., 2008]. Majority of work emphasizes $O(n) \geq O(p)$.

1. Convergence is studied from the perspective of matrix norms.
2. Convergence speed is theorized in the $O(\cdot)$-style, emphasizing how increasing samples, dimensions, and factors influence error.
3. Simulations demonstrate adequacy of a proposed norm.
4. Analysis is described as 'theoretical'

Factor model review

Our n data, Y, are in p dimensions with k factors.

1. $\Psi \in \mathbb{R}_{\geq 0}^{p \times p}$, *diagonal*, *not random*; $L = [l_{i,j}] \in \mathbb{R}^{p \times k}$, *not random*
2. $F = [f_j] \sim N(0, I_k); E = [e_i] \sim N(0, \Psi)$
3. E does not depend on F
4. $Y = \mu + LF + E \Rightarrow Y \sim N(\mu, \Sigma)$ $s.t.$ $Cov[Y] = \Sigma = LL' + \Psi$

Estimation for L can be done in a few ways, of which MLE and Eigenvectors are popular. The authors use least squares given Y & F.

1. $\hat{\Sigma} = \hat{B}\hat{Cov}[F]\hat{B}' + \hat{\Psi}, \hat{\Psi} = diag[E_v E_v'/n], \hat{B} = YX'(XX')^{-1}$
2. such that $E_v = [E_1, \ldots, E_n], X = [F_1, \ldots, F_n]$
3. Criticism: I don't get to observe F! So it is a theoretical exercise.

Matrix norms

This paper primarily uses the following norms.

1. Frobenious [Horn & Johnson, 1990]:
   $||A|| = [tr(AA')]^{1/2} = [\sum_{j=1}^{p} \lambda(A)^2]^{1/2}$
2. Entropy loss: [James & Stein, 1961]
   $L_1(\hat{\Sigma}, \Sigma) = tr\left(\hat{\Sigma}\Sigma^{-1}\right) - \log\left|\hat{\Sigma}\Sigma^{-1}\right| - p$
3. Custom $||A||_\Sigma = \sqrt{p}^{-1}||\Sigma^{-1/2}A\Sigma^{-1/2}||$

Their custom norm is a scaled quadratic norm L_2. Precisely,
$||A - \Sigma||_\Sigma = \sqrt{p}^{-1}L_2(\hat{\Sigma}, \Sigma) = \sqrt{p}^{-1}tr\left[A\Sigma^{-1} - I\right]$.
Convention: Eigenvalues of A are $\lambda_1(A), \ldots, \lambda_p(A)$ decreasing.

Assumptions

The following assumptions are employed partially or entirely.
Let $b = \mathbb{E}||Y||^2, c = \max_{1 \leq i \leq k} \mathbb{E}[f_j^4], d = \max_{1 \leq i \leq k} \mathbb{E}[e_i^4]$

1. (A) (Y_i, F_i) $iid, \mathbb{E}[E_i|F_i] = 0, Cov[E|F] = \Psi$, distribution of F is continuous, $k \leq p$.
2. (B) $b = O(p), c$ & d bounded. $\exists \sigma_1 > 0 \ : \ \lambda_k(LL') \geq \sigma_1 \ \forall n$
3. (C) $\exists \sigma_2 > 0 \ : \ \lambda_p(\Psi) > \sigma_2 \ \forall n$
4. (D) The k factors $F = [f_j]$ are fixed accross n.

Frobenious is a poor descriptor of convergence

Theorem (1)

Under assumptions (A) - (C),

1. $||\hat{\Sigma} - \Sigma|| = O_P(pk/\sqrt{n})$
2. $||\hat{\Sigma}_s - \Sigma|| = O_P(pk/\sqrt{n})$
3. $\max_{1 \leq i \leq p} \left| \lambda_i(\hat{\Sigma}) - \lambda_i(\Sigma) \right| = O_P(pk/\sqrt{n})$
4. $\max_{1 \leq i \leq p} \left| \lambda_i(\hat{\Sigma}_s) - \lambda_i(\Sigma) \right| = O_P(pk/\sqrt{n})$

Notation

1. Let $\hat{\Sigma}$ denote the factor model estimate of Σ.
2. Let $\hat{\Sigma}_s$ denote the sample covariance estimate of Σ.

Quadratic norm better describes convergence

Theorem (2)

Under assumptions (A) - (C), if $K = O(n^{\alpha_1}), p = O(n^{\alpha})$, *then*

1. $||\hat{\Sigma} - \Sigma||_{\Sigma} = O_P(n^{-\beta/2}) \ : \ \beta = \min\{1 - 2\alpha_1, 2 - \alpha - \alpha_1\}$
2. $||\hat{\Sigma}_s - \Sigma||_{\Sigma} = O_P(n^{-\beta_1/2}) \ : \ \beta_1 = 1 - \max\{\alpha, 3\alpha_1/2, 3\alpha_1 - \alpha\}$

Remark: $\alpha > 2\alpha_1$ & $\alpha_1 < 1 \Rightarrow \beta > \beta_1$
So the factor estimate converges faster when the number of dimensions grows faster than the number of factors.
Also note that $\alpha \leq 1 \Rightarrow \hat{\Sigma}$ is root-n-consistent under $|| \cdot ||_{\Sigma}$.

More theory

Theorem (3)

Under assumptions (A) - (C),

1. $||\hat{\Sigma}^{-1} - \Sigma^{-1}|| = o_P([p^2 k^4 \log n/n]^{1/2})$
2. $||\hat{\Sigma}_s^{-1} - \Sigma^{-1}|| = o_P([p^4 k^2 \log n/n]^{1/2})$

Remark: $\hat{\Sigma}^{-1}$ & $\hat{\Sigma}_s^{-1}$ do refer matrix inverses of the estimates.

Theorem (4)

Under assumptions (A), (B), and (D), $\hat{\Sigma} - \Sigma$ converges in distribution to normal errors.
Language is purposefully vague for brevity.

Markowitz portfolio optimization

Define the following variables.

1. There are p risky assets with rate of return μ and covariance Σ.
2. ξ is the variance-minimum allocation vector with rate of return γ.
3. ξ_g is the minimum variance porfolio (no desired rate of return).

Theorem (Markowitz [Markowitz, H.M., 1952])

The solution to the problem:

$$\min_{\xi} \xi' \Sigma \xi \;:\; \xi' \mathbf{1} = 1 \;\&\; \xi \mu = \gamma$$

is this: $\xi = \frac{\phi - \gamma\psi}{\rho\phi - \psi^2} \Sigma^{-1} \mathbf{1} + \frac{\gamma\rho - \psi}{\rho\phi - \psi^2} \Sigma^{-1} \mu$
such that $\rho = \mathbf{1}' \Sigma^{-1} \mathbf{1}, \psi = \mathbf{1}' \Sigma^{-1} \mu, \phi = \mu' \Sigma^{-1} \mu$
with variance $\xi' \Sigma \xi = \frac{\rho\gamma^2 - 2\phi\gamma + \phi}{\rho\phi - \psi^2}$.
To get $\hat{\xi}_g$, *replace* γ *with* ψ / ϕ.

Factor estimates have less porfolio variance

If we simply plug in estimates $\hat{\Sigma}, \hat{\Sigma}_n, \hat{\mu}$ we get allocation estimates $\hat{\xi}, \hat{\xi}_g$.

Theorem (5, global minimum variance convergence)

Under (A) - (C) and $\rho > 0$,

1. $\hat{\xi}_g' \hat{\Sigma} \hat{\xi}_g - \xi_g' \Sigma \xi_g = o_P([p^4 k^4 \log n / n]^{1/2})$
2. $\hat{\xi}_g' \hat{\Sigma}_s \hat{\xi}_g - \xi_g' \Sigma \xi_g = o_P([p^6 k^2 \log n / n]^{1/2})$

Theorem (5, optimal variance convergence)

Under (A) - (C) and
$\rho\phi - \psi^2 > \max\{0, \rho/(\rho\phi - \psi^2), \psi/(\rho\phi - \psi^2), \phi/(\rho\phi - \psi^2)\}$,

1. $\hat{\xi}' \hat{\Sigma} \hat{\xi} - \xi' \Sigma \xi = o_P([p^4 k^4 \log n / n]^{1/2})$
2. $\hat{\xi}' \hat{\Sigma}_s \hat{\xi} - \xi' \Sigma \xi = o_P([p^6 k^2 \log n / n]^{1/2})$

Weak convergence of variance

Require $\xi = O(1)\mathbf{1}$ to avoid extreme short positions.

Theorem

Under (A) and (B),

1. $\xi'\hat{\Sigma}\xi - \xi'\Sigma\xi = o_P([p^4 k^2 \log n/n]^{1/2})$
2. $\xi'\hat{\Sigma}_s\xi - \xi'\Sigma\xi = o_P([p^4 k^2 \log n/n]^{1/2})$

and requiring NO short positions results in

1. $\xi'\hat{\Sigma}\xi - \xi'\Sigma\xi = o_P([p^2 k^2 \log n/n]^{1/2})$
2. $\xi'\hat{\Sigma}_s\xi - \xi'\Sigma\xi = o_P([p^2 k^2 \log n/n]^{1/2})$

Notice the equivalence.

A simulation study

Under fixed n and k, p was allowed to increase. The study revealed the following.

1. The Frobenious norm was insufficient to demonstrate covariance estimate inefficiency.
2. The scaled quadratic norm was sufficient to demonstrate covariance estimate efficiency.
3. When data follow a factor model, observing the factors and including them in the model greatly increases covariance modelling efficiency
4. Estimated portfolio variance MSEs greatly reduce with the factor model when data follow a factor model, even when $p > n$.

Recap

1. Estimation is easy when the model is known and latent variables are observable.
2. Results suggest that correct usage of a factor model may result in reduce variance in covariance estimation and porfolio variance.
3. Results suggest that a simple porfolio ($\xi = O(1)\mathbf{1}$) without shorts does not enjoy a better variance reduction, so standard covariance estimation is adequate.
4. The Frobenious norm may is not always adequate for understanding covariance estimate inefficiency.

References

Jianquing Fan, Yingying Fan, Jinchi Lv (2008)
High dimensional covariance matrix estimation using a factor model
Journal of Econometrics 147, 186-197.

Horn, R.A. & Johnson, C.R. (1990)
Matrix Analysis Cambridge Press

James, W. & Stein, C. (1961)
Estimation iwth quadratic loss
Proc. Fourth Berkeley Symp. Math. Stat. Prob. vol 1, California Press, Berkeley, pp. 361-379

Markowitz (1952)
Portfolio selection
Journal of Finance 7, 77-91.

The End