Martingale approximation for common factor representation

Bystrov, Victor and di Salvatore, Antonietta

University of Lodz, University of Rome "La Sapienza"

26 March 2012

Martingale Approximation of Eigenvalues for Common Factor Representation

Victor Bystrov*, Antonietta di Salvatore†

Abstract

In this paper a martingale approximation is used to derive the limiting distribution of simple positive eigenvalues of the sample covariance matrix for a stationary linear process. The derived distribution can be used to study stability of the common factor representation based on the principal component analysis of the covariance matrix.

Keywords: martingale approximation, dynamic factor model, eigenvalue, stability

1 Introduction

The common factor representation of multivariate time series, which is based on the principal component analysis, is extensively used in the economic forecasting. For the accurate forecasting, it is important to have stable loadings of common factors onto individual time series (Banerjee et al, 2009; Stock and Watson, 2009). The matrix of loadings is estimated by eigenvectors of the sample covariance matrix. As the solution of the eigenvector problem is a pair composed of an eigenvector and an eigenvalue, the problem of studying stability of an eigenvector can be reduced to the problem of studying stability of the corresponding eigenvalue.

In the asymptotic theory, developed for the dynamic factor models (Stock and Watson, 1999; Bai 2003), the limiting distributions of common factors and their loadings are derived under the assumption that both time series and cross section dimensions are increasing. It is not the case in a typical forecasting application where the cross-section dimension is large but fixed, and the time series dimension is increasing.

*University of Lodz, e-mail: emfvib@uni.lodz.pl

†University of Rome, La Sapienza, e-mail: antonietta.disalvatore@uniroma1.it

In this paper we do not rely on the asymptotics developed for dynamic factor models, but assume that a multivariate time series is generated by a linear stationary process of a fixed cross-section dimension, and the factor model is considered as a representation of this time series. We use a martingale approximation of partial sums to derive the asymptotic distribution of simple positive eigenvalues of the sample covariance matrix. This distribution can be used to study stability of the common factor representation.

2 Model Setup

Consider an N-dimensional process $\{X_t\}$ that admits an infinite moving-average representation,

$$X_t = \sum_{i=0}^{\infty} B_i \varepsilon_{t-i},$$

where

Assumption 1: $\{\varepsilon_t\}$ is an N-variate independent identically distributed (i.i.d.) sequence, $E[\varepsilon_0] = 0$, $E|\varepsilon_{0i}\varepsilon_{0j}\varepsilon_{0k}\varepsilon_{0l}| < \infty$ for any $i, j, k, l = 1, 2, ..., N$;

Assumption 2: $\sum_{s=1}^{\infty} s\|B_s\|^2 < \infty$.

We use $\|\cdot\|$ to denote Euclidian norm for vectors and the induced spectral norm for matrices. Let us denote $\Sigma_\varepsilon = E[\varepsilon_0\varepsilon_0']$ and $\Sigma_{\varepsilon\varepsilon} = E[(\varepsilon_0\varepsilon_0' - \Sigma_\varepsilon) \otimes (\varepsilon_0\varepsilon_0' - \Sigma_\varepsilon)]$. Assumption 1 implies that $\|\Sigma_\varepsilon\| < \infty$, and $\|\Sigma_{\varepsilon\varepsilon}\| < \infty$.

Assumptions 1 and 2 imply stationarity and linearity of the process $\{X_t\}$. Though the asymptotic theory, developed for dynamic factor models, does not require neither stationarity nor linearity of common factors, non-stationary time series are usually transformed to make them stationary before applying the principal component analysis, and common factors are often modeled as a vector autoregression.

Let us define the covariance matrix $\Sigma_X = E[X_0X_0'] = \sum_{s=0}^{\infty} B_i\Sigma_\varepsilon B_i'$.

Assumption 3: $\gamma_1 > \gamma_2 > ... > \gamma_r > 0$ ($r \leq N$) are simple positive eigenvalues of matrix Σ_X and $\lambda_1, \lambda_2, ..., \lambda_r$ are corresponding orthonormal eigenvectors.

The assumption of simple eigenvalues is not restrictive for the purpose of empirical analysis, as for a continuous probability distribution, the eigenvalues of the estimated covariance matrix are simple with probability one in a finite sample.

Suppose that a series of T observations is available: $X_1, X_2, ..., X_T$. Consider the estimator $\widehat{\Sigma}_X^{(T)} = T^{-1}\sum_{t=1}^{T} X_t X_t'$ of the matrix Σ_X. Let $\widehat{\gamma}_1^{(T)} > \widehat{\gamma}_2^{(T)} > ... > \widehat{\gamma}_r^{(T)} > 0$ be r largest eigenvalues of $\widehat{\Sigma}_X^{(T)}$, and $\widehat{\lambda}_1^{(T)}, \widehat{\lambda}_2^{(T)}, ..., \widehat{\lambda}_r^{(T)}$ be corresponding orthonormal eigenvectors. Given these estimates, we can consider the decomposition

$$\widehat{\Sigma}_X^{(T)} = \widehat{\Lambda}^{(T)}\widehat{\Gamma}^{(T)}\widehat{\Lambda}^{(T)\prime} + \widehat{\Upsilon}^{(T)},$$

where $\widehat{\Gamma}^{(T)}$ is an $(r \times r)$ diagonal matrix with $(\widehat{\gamma}_1^{(T)}, \widehat{\gamma}_2^{(T)}, ..., \widehat{\gamma}_r^{(T)})$ at the main diagonal, $\widehat{\Lambda}^{(T)} = (\widehat{\lambda}_1^{(T)}, \widehat{\lambda}_2^{(T)}, ..., \widehat{\lambda}_r^{(T)})$, and $\widehat{\Upsilon}^{(T)}$ is an $(r \times r)$ residual matrix. Then the common factor representation of X_t $(t = 1, 2, ..., T)$ is

$$X_t = \widehat{\Lambda}^{(T)}\widehat{F}_t^{(T)} + \widehat{\upsilon}_t^{(T)}, \tag{1}$$

where $\widehat{F}_t^{(T)} = \widehat{\Lambda}^{(T)\prime}X_t$ is an $(r \times 1)$ vector of estimated common factors such that $T^{-1}\sum_{t=1}^{T}\widehat{F}_t^{(T)}\widehat{F}_t^{(T)\prime} = \widehat{\Lambda}^{(T)\prime}(T^{-1}\sum_{t=1}^{T} X_t X_t')\widehat{\Lambda}^{(T)} = \widehat{\Gamma}^{(T)}$, $\widehat{\Lambda}^{(T)}$ is the matrix of loadings of the estimated factors onto observed variables, and $\widehat{\upsilon}_t^{(T)}$ is a vector of residuals.

3 Preliminary Results

3.1 Vectorization

Consider the outer-product

$$X_t X_t' = \left(\sum_{i=0}^{\infty} B_i \varepsilon_{t-i}\right)\left(\sum_{i=0}^{\infty} B_i \varepsilon_{t-i}\right)' = A_t + C_t + C_t',$$

where $A_t = \sum_{i=0}^{\infty} B_i \varepsilon_{t-i}\varepsilon_{t-i}' B_i'$ and $C_t = \sum_{j=1}^{\infty}\sum_{i=0}^{\infty} B_i \left(\varepsilon_{t-i}\varepsilon_{t-i-j}'\right) B_{i+j}'$. For the symmetric matrix $(C_t + C_t')$, it holds that $vec(C_t + C_t') = 2P_N vec(C_t)$, where $P_N = D_N(D_N'D_N)^{-1}D_N'$ and

D_N is a duplication matrix. Let $Y_t = vec(X_t X_t')$. Then

$$Y_t = vec(A_t) + vec(C_t + C_t') = vec(A_t) + 2P_N vec(C_t), \tag{2}$$

where $vec(A_t) = \sum_{i=0}^{\infty} (B_i \otimes B_i)\, vec(\varepsilon_{t-i}\varepsilon_{t-i}')$ and $vec(C_t) = \sum_{j=1}^{\infty}\sum_{i=0}^{\infty} (B_i \otimes B_{i+j})\, vec(\varepsilon_{t-i}\varepsilon_{t-i-j}')$.

3.2 Beveridge-Nelson decomposition

Equation (2) can be rewritten using lag polynomials,

$$Y_t \;\; = F_0(L) vec(\varepsilon_t\varepsilon_t') + 2P_N \textstyle\sum_{j=1}^{\infty} F_j(L) vec(\varepsilon_t\varepsilon_{t-j}'), \tag{3}$$

where $F_0(L) = \sum_{i=0}^{\infty} (B_i \otimes B_i)\, L^i$ and $F_k(L) = \sum_{i=0}^{\infty} (B_i \otimes B_{i+k})\, L^i$. The multivariate Beveridge-Nelson decomposition can be applied to the polynomial $F_k(L)$ giving

$$F_k(L) = F_k(1) - (1-L)\tilde{F}_k(L), \quad \text{where} \quad \tilde{F}_k(L) = \textstyle\sum_{l=0}^{\infty}\left(\sum_{m=l+1}^{\infty} B_m \otimes B_{m+k}\right) L^l.$$

Then it is possible to rewrite equation (3) as

$$Y_t = (F_0(1) - (1-L)\tilde{F}_0(L)) vec(\varepsilon_t\varepsilon_t') + 2P_N \sum_{j=1}^{\infty} (F_j(1) - (1-L)\tilde{F}_j(L)) vec(\varepsilon_t\varepsilon_{t-j}').$$

After rearrangement,

$$\begin{aligned} Y_t = \;\; & F_0(1) vec(\varepsilon_t\varepsilon_t') + 2P_N \textstyle\sum_{j=1}^{\infty} F_j(1) vec(\varepsilon_t\varepsilon_{t-j}') + \\ & -(1-L)\tilde{F}_0(L) vec(\varepsilon_t\varepsilon_t') - 2P_N(1-L) \textstyle\sum_{j=1}^{\infty} \tilde{F}_j(L) vec(\varepsilon_t\varepsilon_{t-j}'). \end{aligned}$$

3.3 Martingale approximation

We can write $vec(\Sigma_X) = F_0(1) vec(\Sigma_\epsilon)$. Let us define

$$Z_t = F_0(1)\left(vec(\varepsilon_t\varepsilon_t') - vec(\Sigma_\epsilon)\right) + 2P_N \sum_{j=1}^{\infty} F_j(1) vec(\varepsilon_t\varepsilon_{t-j}').$$

It can be easily shown that $\{Z_t\}$ is a martingale difference sequence. Let

$$R_t = (1-L)\tilde{F}_0(L)vec(\varepsilon_t\varepsilon_t') + 2P_N(1-L)\sum_{j=1}^{\infty}\tilde{F}_j(L)vec(\varepsilon_t\varepsilon_{t-j}')$$

be a residual term. Then we have

$$Y_t - vec(\Sigma_X) = Z_t - R_t$$

Lemma 1. Under Assumptions 1-3, $T^{-1}\sum_{t=1}^{T} Z_t Z_t' \rightarrow_{a.s.} \Sigma_Z$ ($\rightarrow_{a.s.}$ stands for almost sure convergence), where

$$\Sigma_Z = F_0(1)\Sigma_{\varepsilon\varepsilon}F_0(1)' + 4P_N\left[\sum_{i=1}^{\infty}F_i(1)(\Sigma_\varepsilon\otimes\Sigma_\varepsilon)F_i(1)'\right]P_N', \tag{4}$$

$rank(\Sigma_Z)\geq r^2$ and $\|\Sigma_Z\|<\infty$. (For the proof, see Appendix).

Let us Define $S_{Yt}=\sum_{s=1}^{t}(Y_s - vec\Sigma_X)$, $S_{Zt}=\sum_{s=1}^{t}Z_s$, and $S_{Rt}=\sum_{s=1}^{t}R_s$. Then we have

$$S_{Yt} = S_{Zt} - S_{Rt}, \tag{5}$$

where $\{S_{Zt}\}$ is a martingale, and $\{S_{Rt}\}$ is a residual sequence.

As $\{Z_t\}$ is a martingale difference sequence and $\{S_{Zt}\}$ is the corresponding martingale, we have $E[T^{-1}S_{ZT}S_{ZT}'] = E[T^{-1}\sum_{t=1}^{T}Z_tZ_t'] = T^{-1}\sum_{t=1}^{T}E[Z_tZ_t']$ and $T^{-1}S_{ZT}S_{ZT}'\rightarrow_{a.s.}\Sigma_Z$.

Lemma 2. Under Assumptions 1 and 2, $T^{-1}S_{YT}\rightarrow_{a.s.} 0$.

The proof of Lemma 2 makes use of decomposition (5) (see Appendix). Lemma 2 implies that Σ_X can be consistently estimated by $T^{-1}\sum_1^T X_tX_t'$, as $T^{-1}S_{YT} = T^{-1}\sum_{t=1}^{T}vec(X_tX_t') - vec\Sigma_X$ converges to zero a.s.

4 Main Results

Consider

$$\gamma = \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_r \end{pmatrix},\ \Lambda_\otimes = \begin{pmatrix} \lambda_1' \otimes \lambda_1' \\ \lambda_2' \otimes \lambda_2' \\ \vdots \\ \lambda_r' \otimes \lambda_r' \end{pmatrix},\ \widehat{\gamma}^{(T)} = \begin{pmatrix} \widehat{\gamma}_1^{(T)} \\ \widehat{\gamma}_2^{(T)} \\ \vdots \\ \widehat{\gamma}_r^{(T)} \end{pmatrix},\ \text{and } \widehat{\Lambda}_\otimes^{(T)} = \begin{pmatrix} \widehat{\lambda}_1^{(T)\prime} \otimes \widehat{\lambda}_1^{(T)\prime} \\ \widehat{\lambda}_2^{(T)\prime} \otimes \widehat{\lambda}_2^{(T)\prime} \\ \vdots \\ \widehat{\lambda}_r^{(T)\prime} \otimes \widehat{\lambda}_r^{(T)\prime} \end{pmatrix}$$

Proposition 1 (Consistency) Under Assumptions 1 - 3, $\widehat{\gamma}^{(T)} \to_{a.s.} \gamma$ and $\widehat{\Lambda}_\otimes^{(T)} \to_{a.s.} \Lambda_\otimes$
The proof immediately follows from Lemma 2 and the continuous mapping theorem, as eigenvectors and eigenvalues are continuous functions of matrix entries.

Using the eigenvalue derivative (Magnus, 1985) and the Taylor expansion (Fuller, 1976, p. 192), we obtain

$$\widehat{\gamma}^{(T)} - \gamma = \Lambda_\otimes vec(\widehat{\Sigma}_X^{(T)} - \Sigma_X) + O_P(T^{-2})$$

where $vec(\widehat{\Sigma}_X^{(T)} - \Sigma_X) = T^{-1}S_{YT}$. As $S_{YT} = S_{ZT} - S_{RT}$, we get

$$T^{1/2}(\widehat{\gamma}^{(T)} - \gamma) = T^{-1/2}\Lambda_\otimes S_{ZT} - T^{-1/2}\Lambda_\otimes S_{RT} + O_P(T^{-3/2})$$

Proposition 2 (Central Limit Theorem). Under Assumptions 1 - 3,

$$T^{1/2}(\widehat{\gamma}^{(T)} - \gamma) \to_d N(0, \Lambda_\otimes \Sigma_Z \Lambda_\otimes').$$

(For the proof, see Appendix).

Given a consistent estimator of $\Lambda_\otimes \Sigma_Z \Lambda_\otimes'$, it is possible to construct recursive confidence intervals for eigenvalues $(\gamma_1, \gamma_2, ..., \gamma_r)' = \gamma$, and analize stability of these eigenvalues and of the corresponding eigenvectors, $(\lambda_1, \lambda_2, ..., \lambda_r) = \Lambda$. If these eigenvectors are stable, so should be their estimates, representing the matrix of loadings $\widehat{\Lambda}^{(T)} = (\widehat{\lambda}_1^{(T)}, \widehat{\lambda}_2^{(T)}, ..., \widehat{\lambda}_r^{(T)})$ in the common factor representation (1).

5 Concluding Remarks

In this paper a martingale approximation is used to derive the limiting distribution of simple positive eigenvalues of the sample covariance matrix for a stationary linear process. The eigenvalues jointly with the corresponding eigenvectors represent a solution of the eigenvector problem. Their limiting distribution can be used for studying stability of this solution, which is equivalent to studying stability of the common factor representation based on the principal component analysis of the covariance matrix. The development of a statistical procedure for the stability analysis is left for future research.

6 References

Bai J., 2003. Inferential theory for factor models of large dimensions, Econometrica, 7(1), 135-171.

Banerjee, A., Marcellino, M., and Masten I., 2009. Forecasting macroeconomic variables using diffusion indexes in short samples with structural change, in D. E. Rapach, M. E. Wohar (ed.) Forecasting in the Presence of Structural Breaks and Model Uncertainty (Frontiers of Economics and Globalization, Volume 3), Emerald Group Publishing Limited, pp.149-194

Fuller, W.A., 1976. Introduction to Statistical Time Series, John Wiley, New York

Heyde, C.C. 1997. Quasi-Likelihood and Its Applications: A General Approach to Optimal Parameter Estimation, Springer-Verlag, New York.

Magnus, J.R., 1985. On differentiating eigenvalues and eigenvectors, Econometric Theory, 1, 179-191.

Stock, J.H., and Watson, M., 1998. Diffusion Indexes, Working Paper 6702, National Bureau of Economic Research.

Stock, J. H., and Watson, M., 2009. Forecasting in dynamic factor models subject to structural instability, in The Methodology and Practice of Econometrics, A Festschrift in Honour of David

F. Hendry, eds. J. Castle and N. Shephard, OUP Oxford

Appendix. Outlines of Proofs

Proof of Lemma 1. Using Assumptions 1 and 2, it can be shown that $\{Z_t Z_t'\}$ is a stationary integrable sequence and $E[Z_t Z_t'] = \Sigma_Z$, where Σ_Z is given by (4). The matrix Σ_Z is finite, as Σ_ε and $\Sigma_{\varepsilon\varepsilon}$ are finite by Assumption 1, and $F_0(1)F_0(1)'$ and $\sum_{i=1}^{\infty} F_i(1)F_i(1)'$ are finite by Assumption 2. By the pointwise ergodic theorem, $T^{-1}\sum_{t=1}^{T} Z_t Z_t' \to_{a.s.} \Sigma_Z$.

Assumption 3 implies that $\Sigma_X = \sum_{s=0}^{\infty} B_i \Sigma_\varepsilon B_i'$ has a rank of at least r. Then the rank of Σ_ε is at least r and the rank of $\Sigma_{\varepsilon\varepsilon}$ is at least r^2. Using decomposition (4), it can be shown that $rank(\Sigma_Z) \geq r^2$.

Proof of Lemma 2. Consider the decomposition $S_{Yt} = S_{Zt} - S_{Rt}$. For the first component, S_{Zt}, we have

$$S_{Zt} = \sum_{s=1}^{t} Z_s = \sum_{s=1}^{t} \left[F_0(1)\left(vec(\varepsilon_s \varepsilon_s') - vec\Sigma_\epsilon\right) + 2P_N \sum_{j=1}^{\infty} F_j(1) vec(\varepsilon_s \varepsilon_{s-j}') \right].$$

The sequence $\{S_{Zt}\}$ is a uniformly L_2-bounded martingale, that satisfies the conditions of Theorem 12.4 in Heyde (1997, p.187), as $\sum_{t=1}^{\infty} E\|t^{-1}Z_t\|^2 = \sum_{t=1}^{\infty} t^{-2} tr(E[Z_t Z_t']) = \sum_{t=1}^{\infty} t^{-2} tr(\Sigma_Z) < \infty$. Then $T^{-1}S_{ZT}$ converges to zero almost surely. The residual S_{Rt} is a telescoping sum of random variables that can be written as

$$S_{Rt} = \tilde{F}_0(L)\left[vec(\varepsilon_t \varepsilon_t') - vec(\varepsilon_0 \varepsilon_0') \right] + 2P_N \sum_{j=1}^{\infty} \tilde{F}_j(L) \left[vec(\varepsilon_t \varepsilon_{t-j}') - vec(\varepsilon_0 \varepsilon_{0-j}') \right].$$

Under Assumptions 1 and 2, $E[S_{Rt}] = 0$ and $E\|S_{Rt}\|^2 < \infty$. Then, using Chebyshev inequality and Borel-Cantelli lemma it can be proven that $T^{-1}S_{RT} \to_{a.s.} 0$.

Proof of Proposition 2. Given the decomposition

$$t^{1/2}(\widehat{\gamma}^{(t)} - \gamma) = t^{-1/2}\Lambda_{\otimes} S_{Zt} - t^{-1/2}\Lambda_{\otimes} S_{RT} + O_P(t^{-3/2}),$$

consider its first component, $\Lambda_{\otimes} S_{Zt}$, which is a martingale satisfying

(i) $t^{-1/2} \sup_{s \leq t} |\Lambda_{\otimes} Z_s| \to_p 0$, where $Z_s = S_{Zs} - S_{Zs-1}$;

(ii) $t^{-1} \sum_{s=1}^{t} \Lambda_{\otimes} Z_s Z_s' \Lambda_{\otimes}' \to_p \Lambda_{\otimes} \Sigma_Z \Lambda_{\otimes}'$;

(iii) $t^{-1} E[\Lambda_{\otimes} S_{Zt} S_{Zt}' \Lambda_{\otimes}'] = \Lambda_{\otimes} \Sigma_Z \Lambda_{\otimes}'$.

Lemma 1 implies that $rank(\Sigma_Z) \geq r^2$. Under Assumption 3, $rank(\Lambda_{\otimes}) = r$ and $\Lambda_{\otimes} \Sigma_Z \Lambda_{\otimes}'$ is positive definite. Then the conditions of Theorem 12.6 in Heyde (1997, p.192) are satisfied. It follows that $T^{-1/2} \Lambda_{\otimes} S_{ZT} \to N(0, \Lambda_{\otimes} \Sigma_Z \Lambda_{\otimes}')$. Using Chebyshev inequality, it is easy to show that the residual term, $T^{-1/2} \Lambda_{\otimes} S_{RT}$, converges to zero in probability, as $E[S_{RT}] = 0$ and $E\|S_{RT}\|^2 < \infty$.