Martingale approximation for common factor representation

Victor Bystrov and Antonietta di Salvatore

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, $E[\varepsilon_0] = 0$, $E[\varepsilon_0 \varepsilon_{0j} \varepsilon_{0k} \varepsilon_{0l}] < \infty$ 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 $\Sigma_X = E[X_0 X_0'] = \sum_{s=0}^{\infty} B_s \Sigma_\varepsilon B_s'$.

**Assumption 3:** $\gamma_1 > \gamma_2 > ... > \gamma_r > 0$ ($r \leq N$) are simple positive eigenvalues of matrix $\Sigma_X$ 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: \( X_1, X_2, \ldots, X_T \). Consider the estimator \( \hat{\Sigma}_X(T) = T^{-1} \sum_{t=1}^{T} X_t X_t' \) of the matrix \( \Sigma_X \). Let \( \hat{\gamma}_1(T) > \hat{\gamma}_2(T) > \ldots > \hat{\gamma}_r(T) > 0 \) be \( r \) largest eigenvalues of \( \hat{\Sigma}_X(T) \), and \( \hat{\lambda}_1(T), \hat{\lambda}_2(T), \ldots, \hat{\lambda}_r(T) \) be corresponding orthonormal eigenvectors. Given these estimates, we can consider the decomposition

\[
\hat{\Sigma}_X(T) = \hat{\Lambda}(T) \hat{\Gamma}(T) \hat{\Lambda}(T)' + \hat{\Upsilon}(T),
\]

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

\[
X_t = \hat{\Lambda}(T) \hat{F}_t(T) + \hat{u}_t(T),
\]

where \( \hat{F}_t = \hat{\Lambda}(T)'X_t \) is an \((r \times 1)\) vector of estimated common factors such that \( T^{-1} \sum_{t=1}^{T} \hat{F}_t(T) \hat{F}_t(T)' = \hat{\Lambda}(T)'(T^{-1} \sum_{t=1}^{T} X_t X_t') \hat{\Lambda}(T) = \hat{\Gamma}(T) \), \( \hat{\Lambda}(T) \) is the matrix of loadings of the estimated factors onto observed variables, and \( \hat{u}_t(T) \) 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 \( A_t = \sum_{i=0}^{\infty} B_i \varepsilon_{t-i} \varepsilon_{t-i}' \) and \( C_t = \sum_{j=1}^{\infty} \sum_{i=0}^{\infty} B_i (\varepsilon_{t-i} \varepsilon_{t-i-j}' B_{i+j} \varepsilon_{t-i-j}' \varepsilon_{t-i-j} \varepsilon_{t-i-j}'). \) 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 $Y_t = vec(X_tX_t')$. Then

$$Y_t = vec(A_t) + vec(C_t + C_t') = vec(A_t) + 2P_N vec(C_t), \quad (2)$$

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

### 3.2 Beveridge-Nelson decomposition

Equation (2) can be rewritten using lag polynomials,

$$Y_t = F_0(L)vec(\varepsilon_{t}\varepsilon'_{t}) + 2P_N \sum_{j=1}^{\infty} F_j(L)vec(\varepsilon_{t}\varepsilon'_{t-j}), \quad (3)$$

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

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

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,

$$Y_t = F_0(1)vec(\varepsilon_{t}\varepsilon'_{t}) + 2P_N \sum_{j=1}^{\infty} F_j(1)vec(\varepsilon_{t}\varepsilon'_{t-j}) +$$

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

### 3.3 Martingale approximation

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

$$Z_t = F_0(1) (vec(\varepsilon_{t}\varepsilon'_{t}) - vec(\Sigma_\varepsilon)) + 2P_N \sum_{j=1}^{\infty} F_j(1)vec(\varepsilon_{t}\varepsilon'_{t-j}).$$
It can be easily shown that \( \{Z_t\} \) is a martingale difference sequence. Let

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

be a residual term. Then we have

\[
Y_t - \text{vec}(\Sigma_X) = Z_t - R_t
\]

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

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

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

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

\[
S_{Yt} = S_{Zt} - S_{Rt},
\]

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 \( T^{-1}S_{ZT}S_{ZT}' \to_{a.s} \Sigma_Z \).

**Lemma 2.** Under Assumptions 1 and 2, \( T^{-1}S_{YT} \to_{a.s} 0 \).

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

### 4 Main Results

Consider
\[ \gamma = \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_r \end{pmatrix}, \quad \Lambda_\otimes = \begin{pmatrix} \lambda_1 \otimes \lambda'_1 \\ \lambda_2 \otimes \lambda'_2 \\ \vdots \\ \lambda_r \otimes \lambda'_r \end{pmatrix}, \quad \hat{\gamma}^{(T)} = \begin{pmatrix} \hat{\gamma}_1^{(T)} \\ \hat{\gamma}_2^{(T)} \\ \vdots \\ \hat{\gamma}_r^{(T)} \end{pmatrix}, \quad \text{and} \quad \hat{\Lambda}_\otimes^{(T)} = \begin{pmatrix} \hat{\lambda}_1^{(T)} \otimes \hat{\lambda}'_1^{(T)} \\ \hat{\lambda}_2^{(T)} \otimes \hat{\lambda}'_2^{(T)} \\ \vdots \\ \hat{\lambda}_r^{(T)} \otimes \hat{\lambda}'_r^{(T)} \end{pmatrix} \]

**Proposition 1 (Consistency)** Under Assumptions 1 - 3, \( \hat{\gamma}^{(T)} \rightarrow_{a.s.} \gamma \) and \( \hat{\Lambda}_\otimes^{(T)} \rightarrow_{a.s.} \Lambda_\otimes \)

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

\[ \hat{\gamma}^{(T)} - \gamma = \Lambda_\otimes vec(\hat{\Sigma}_X^{(T)} - \Sigma_X) + O_P(T^{-2}) \]

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

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

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

\[ T^{1/2}(\hat{\gamma}^{(T)} - \gamma) \rightarrow_d N(0, \Lambda_\otimes \Sigma Z \Lambda_\otimes') \]

(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 \( (\gamma_1, \gamma_2, \ldots, \gamma_r)' = \gamma \), and analyze stability of these eigenvalues and of the corresponding eigenvectors, \( (\lambda_1, \lambda_2, \ldots, \lambda_r) = \Lambda \). If these eigenvectors are stable, so should be their estimates, representing the matrix of loadings \( \hat{\Lambda}^{(T)} = (\hat{\lambda}_1^{(T)}, \hat{\lambda}_2^{(T)}, \ldots, \hat{\lambda}_r^{(T)}) \) 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

Appendix. Outlines of Proofs

Proof of Lemma 1. Using Assumptions 1 and 2, it can be shown that \( \{Z_tZ'_t\} \) is a stationary integrable sequence and \( E[Z_tZ'_t] = \Sigma_Z \), where \( \Sigma_Z \) is given by (4). The matrix \( \Sigma_Z \) is finite, as \( \Sigma_\epsilon \) and \( \Sigma_{\epsilon\epsilon} \) 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, \( T^{-1} \sum_{t=1}^{T} Z_tZ'_t \to_{a.s.} \Sigma_Z \).

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

Proof of Lemma 2. Consider the decomposition \( S_Y = S_{Zt} - S_{Rt} \). For the first component, \( S_{Zt} \), we have

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

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 \( \sum_{t=1}^{\infty} E[t^{-1}Z_t]^2 = \sum_{t=1}^{\infty} t^{-2}tr(E[Z_tZ'_t]) = \sum_{t=1}^{\infty} t^{-2}tr(\Sigma_Z) < \infty \). 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

\[
S_{Rt} = \tilde{F}_0(L) \left[ \text{vec}(\epsilon_t\epsilon'_t) - \text{vec}(\epsilon_0\epsilon'_0) \right] + 2P_N \sum_{j=1}^{\infty} \tilde{F}_j(L) \left[ \text{vec}(\epsilon_t\epsilon'_{t-j}) - \text{vec}(\epsilon_0\epsilon'_{0-j}) \right].
\]

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

\[
t^{1/2}(\tilde{\gamma}(0) - \gamma) = t^{-1/2} \Lambda_\otimes S_{Zt} - t^{-1/2} \Lambda_\otimes S_{Rt} + O_P(t^{-3/2}),
\]
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) $t^{-1} \sum_{s=1}^{t} \Lambda \otimes Z_s Z_s' \Lambda' \to_p \Lambda \otimes \Sigma Z \Lambda'$;

(iii) $t^{-1} E[\Lambda \otimes S_{Zt} S'_{Zt} \Lambda'] = \Lambda \otimes \Sigma Z \Lambda'$.

Lemma 1 implies that $\text{rank}(\Sigma_Z) \geq r^2$. Under Assumption 3, $\text{rank}(\Lambda) = r$ and $\Lambda \otimes \Sigma Z \Lambda'$ is positive definite. Then the conditions of Theorem 12.6 in Heyde (1997, p.192) are satisfied. It follows that $T^{-1/2} \Lambda \otimes S_{Zt} \to N(0, \Lambda \otimes \Sigma Z \Lambda')$. 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 $E\|S_{RT}\|^2 < \infty$. 