LARGE COVARIANCE MATRICES FOR TIME SERIES DATA

Abstract. A random matrix theory has found an extraordinary number of applications in mathematics, statistics, physics or finance. One of particular problems is to study a behaviour of (largest) eigenvalues of a covariance matrix XX' associated with a data matrix X . The data matrix is defined by p rows and n columns, where p is possibly much larger than the number n . Intuitively, p can represent the number of different stocks and then n describes their prices at distinct time points. The objective of this research is to observe the possible limiting distributions for (largest) eigenvalues of large covariance matrices when independence and normality assumptions are dropped. In this project, we will use existing theoretical results for independent data as a benchmark for our extensive Monte Carlo experiments. We simulate a large number of time series and compute empirical distribution of eigenvalues of the associated sample covariance matrix.

1.1 Definition of eigenvalues of covariance matrices

We consider a data matrix

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \cdot \\ \cdot \\ X_p \end{bmatrix}$$

where

$$X_1 = (X_{11}, X_{12}, \ldots\ldots, X_{1n}),$$
$$X_2 = (X_{21}, X_{22}, \ldots\ldots, X_{2n}),$$
$$\ldots\ldots$$
$$X_p = (X_{P1}, X_{P2}, \ldots\ldots, X_{pn}).$$

Each row in the matrix X is a time series observed at n time points. In practice, usually n and p are large.
We are interested in eigenvalues of the scaled sample covariance matrix :

$$B = \frac{1}{n-1} X \cdot X^T.$$

Such problems arise naturally in statistics (e.g. principal components analysis) or finance (e.g. risk allocation) among others.

2.1 WHAT IS KNOWN?

Assume that each row consists of independent and identically distributed random variables. Furthermore, assume that rows are independent of each other. Then the eigenvalues follow **Marchenko-Pastur Law**:

$$L(dx) = \frac{1}{2\pi x}\sqrt{(x_+ - x)(x - x_-)}$$

where

$$x_- < x < x_+$$

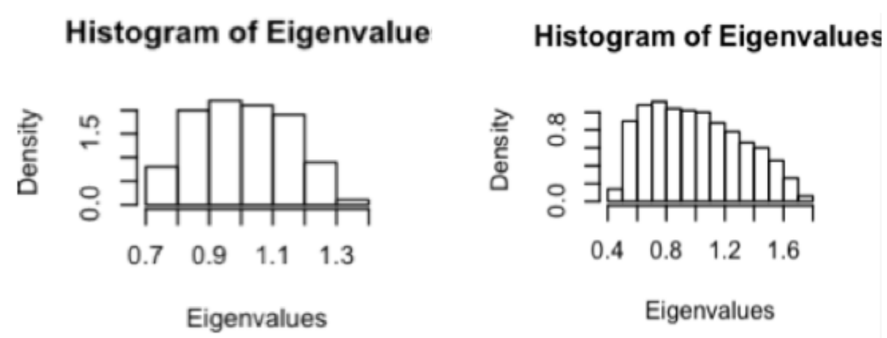
and

$$x_\pm = (1 + \sqrt{y})^2; \qquad y = \lim \frac{p}{n}.$$

2.2 Illustration of Marchenko-Pastur Law

We illustrate the Marchenko- Pastur Law for i.i.d. data as a benchmark for comparison with time series models.

- Graph1: n=5000,p=100;
- Graph2: n=5000,p=500.



CONCLUSIONS:

- When Xi are i.i.d. standard normal, then the empirical distribution of the eigenvalues follows Marcenko-Pasutr Law. This is best illustrated for n = 5000, p = 500; see Graph2.
- Now, consider Xi to be AR(1), in the same case of n = 5000, p = 500. When ρ = 0.5, the histogram has a similar shape as in the i.i.d. case, but with a bigger range of eigenvalues; for ρ=0.8, it has a different shape as compared to the i.i.d. case, and also has a bigger range of eigenvalues. This empirical observations suggests that the Marchenko-Pastur Law may not be valid in time series case.
- At last, consider Xi to be SV, in the same case of n = 5000, p = 500. When ρ = 0.5, the histogram has a different shape as compared to the i.i.d.case, and with a bigger range of eigenvalues; the same for ρ = 0.8. This observation suggests that the Marchenko-Pastur Law may not be valid in stochastic volatility case.

3.1 AR(1) model

AutoRegressive model of order 1 is defined as a solution to a recurrence equation

$$X_i = \rho X_{i-1} + Z_i.$$

where

$$|\rho| < 1.$$

Assume that Zi is a sequence of i.i.d random variables with mean zero.

3.2 SV model

Stochastic volatility (SV) model is defined as

$$X_i = \sigma(Y_i) Z_i,$$

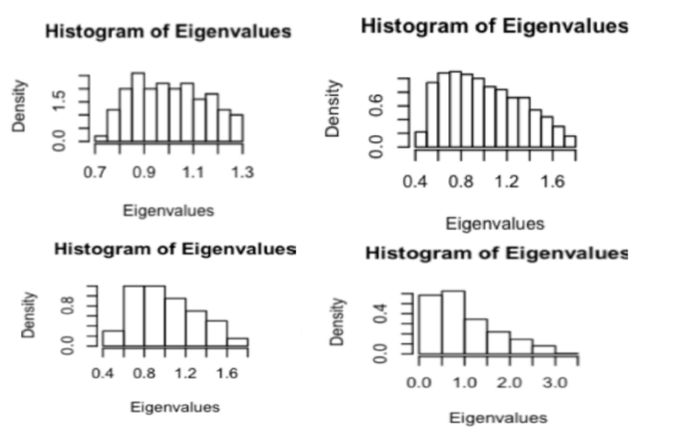
where σ is a positive function, Yi is stationary sequence, such that Yi and Zi are independent for given i.

Question :

How do the eigenvalues behave for different time series models?

Empirical distribution of eigenvalues for AR(1) model

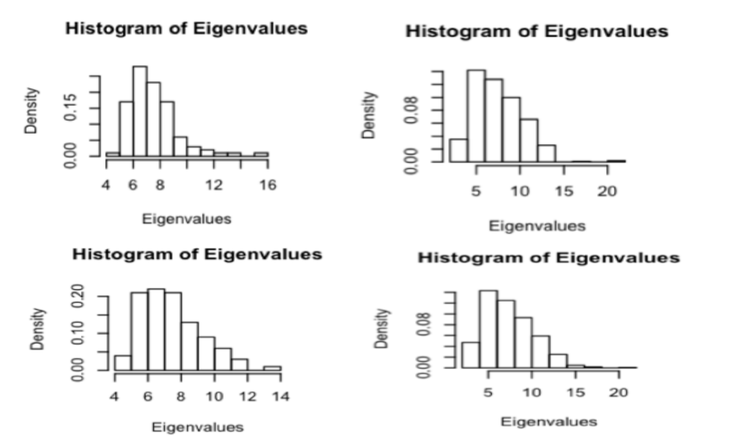
- Graph 3: n = 5000, p = 100, ρ=0.2;
- Graph 4: n = 5000, p = 500, ρ=0.2;
- Graph 5: n = 5000, p = 100, ρ=0.8;
- Graph 6: n = 5000, p = 500, ρ=0.8;



Empirical distribution of eigenvalues for SV model

We choose Yi to be autoregressive model of order 1, with different values of ρ:

- Graph 7: n = 5000, p = 100, ρ=0.2;
- Graph 8 : n = 5000, p = 500, ρ=0.2;
- Graph 9: n = 5000, p = 100, ρ=0.8;
- Graph 10: n = 5000, p = 500, ρ=0.8.



Thank you to Professor Rafal Kulik for his supervision and to the UROP project for this opportunity

WENQIAN LI
wli046@uottawa.ca