# 17

# Sample Covariance Matrices

In this chapter, we will show how to compute the various transforms ($S(t)$, $\mathfrak{t}(z)$, $\mathfrak{g}(z)$) for sample covariance matrices (SCM) when the data has non-trivial true correlations, i.e. is characterized by a non-diagonal true underlying covariance matrix $\mathbf{C}$ and possibly non-trivial temporal correlations as well. More precisely, $N$ time series of length $T$ are stored in a rectangular $N \times T$ matrix $\mathbf{H}$. The *sample covariance matrix* is defined as

$$\mathbf{E} = \frac{1}{T}\mathbf{H}\mathbf{H}^T. \tag{17.1}$$

If the $N$ time series are stationary, we expect that for $T \gg N$, the SCM $\mathbf{E}$ converges to the "true" covariance matrix $\mathbf{C}$. The non-trivial correlations encoded in the off-diagonal elements of $\mathbf{C}$ are what we henceforth call *spatial* (or cross-sectional) correlations. But the $T$ samples might also be non-independent and we will also model these *temporal* correlations. Of course, the data might have both types of correlations (spatial and temporal).

We will be interested in the eigenvalues $\{\lambda_k\}$ of $\mathbf{E}$ and their density $\rho_{\mathbf{E}}(\lambda)$, which we will compute from the knowledge of its Stieltjes transform $\mathfrak{g}_{\mathbf{E}}(z)$ using Eq. (2.47). We can also compute the singular values $\{s_k\}$ of $\mathbf{H}$; note that these singular values are related to the eigenvalues of $\mathbf{E}$ via $s_k = \sqrt{T\lambda_k}$.

## 17.1 Spatial Correlations

Consider the case where $\mathbf{H}$ are multivariate Gaussian observations, drawn from $\mathcal{N}(0, \mathbf{C})$. We saw in Section 4.2.4 that $\mathbf{E}$ is then a general Wishart matrix with column covariance $\mathbf{C}$, and can be written as

$$\mathbf{E} = \mathbf{C}^{\frac{1}{2}}\mathbf{W}_q\mathbf{C}^{\frac{1}{2}}. \tag{17.2}$$

We recognize this formula as the free product of the covariance matrix $\mathbf{C}$ and a white Wishart of parameter $q$, $\mathbf{W}_q$. Note that since the white Wishart is rotationally invariant, it is free from any matrix $\mathbf{C}$. From the multiplicativity of the S-transform and the form of the S-transform of the white Wishart (Eq. (15.21)), we have

$$S_{\mathbf{E}}(t) = \frac{S_{\mathbf{C}}(t)}{1 + qt}. \tag{17.3}$$



https://doi.org/10.1017/9781108768900.021 Published online by Cambridge University Press

We can also use the subordination relation of the free product (Eq. (11.109)) to write this relation in terms of T-transforms:

$$\mathfrak{t}_{\mathbf{E}}(z) = \mathfrak{t}_{\mathbf{C}}\left(Z(z)\right), \qquad Z(z) = \frac{z}{1 + q\mathfrak{t}_{\mathbf{E}}(z)}. \tag{17.4}$$

This last expression can be written in terms of the more familiar Stieltjes transform using $\mathfrak{t}(z) = z\mathfrak{g}(z) - 1$:

$$z\mathfrak{g}_{\mathbf{E}}(z) = Z\mathfrak{g}_{\mathbf{C}}(Z), \quad \text{where} \quad Z = \frac{z}{1 - q + qz\mathfrak{g}_{\mathbf{E}}(z)}. \tag{17.5}$$

This is the central equation that allows one to infer the "true" spectral density of $\mathbf{C}$, $\rho_{\mathbf{C}}(\lambda)$, from the empirically observed spectrum of $\mathbf{E}$. Note that this equation can equivalently be rewritten in terms of the spectral density of $\mathbf{C}$ as

$$\mathfrak{g}_{\mathbf{E}}(z) = \int \frac{\rho_{\mathbf{C}}(\mu)\mathrm{d}\mu}{z - \mu(1 - q + qz\mathfrak{g}_{\mathbf{E}}(z))}. \tag{17.6}$$

We will see in Chapter 20 some real world applications of this formula. One of the most important properties of Eq. (17.5) is its universality: it holds (in the large $N$ limit) much beyond the restricted perimeter of multivariate Gaussian observations $\mathbf{H}$. In fact, as soon as the observations have a finite second moment, the relation between the "true" spectral density $\rho_{\mathbf{C}}$ and the empirical Stieltjes transform $\mathfrak{g}_{\mathbf{E}}(z)$ is given by Eq. (17.5).

Let us discuss some interesting limiting cases. First, when $q \to 0$, i.e. when $T \gg N$, one expects that $\mathbf{E} \approx \mathbf{C}$. This is indeed what one finds since in that limit $Z = z + O(q)$; hence $\mathfrak{g}_{\mathbf{E}}(z) = \mathfrak{g}_{\mathbf{C}}(z)$ and $\rho_{\mathbf{E}} = \rho_{\mathbf{C}}$.

Second, consider the case $\mathbf{C} = \mathbf{1}$, for which $\mathfrak{g}_{\mathbf{C}}(Z) = 1/(Z-1)$. We thus obtain

$$z\mathfrak{g}_{\mathbf{E}}(z) = \frac{Z}{Z-1} = \frac{z}{(z - 1 + q - qz\mathfrak{g}_{\mathbf{E}}(z))} \to \frac{1}{\mathfrak{g}_{\mathbf{E}}(z)} = z1 + q - qz\mathfrak{g}_{\mathbf{E}}(z), \tag{17.7}$$

which coincides with Eq. (4.37). In the next exercise, we consider the case where $\mathbf{C}$ is an inverse-Wishart matrix, in which case some explicit results can be obtained.

We can also infer some properties of the spectrum of $\mathbf{E}$ using the moment generating function. The T-transform of $\mathbf{E}$ can be expressed as the following power series for $z \to \infty$:

$$\mathfrak{t}_{\mathbf{E}}(z) \underset{z\to\infty}{\longrightarrow} \sum_{k=1}^{\infty} \tau(\mathbf{E}^k) z^{-k}. \tag{17.8}$$

We thus deduce that

$$Z(z) \underset{z\to\infty}{\longrightarrow} \frac{z}{1 + q\sum_{k=1}^{\infty}\tau(\mathbf{E}^k)z^{-k}}.$$

Therefore we have, for $z \to \infty$,

$$\mathfrak{t}_{\mathbf{C}}(Z(z)) \underset{z\to\infty}{\longrightarrow} \sum_{k=1}^{\infty} \frac{\tau(\mathbf{C}^k)}{z^k}\left(1 + q\sum_{\ell=1}^{\infty}\tau(\mathbf{E}^{\ell})z^{-\ell}\right)^k. \tag{17.9}$$

https://doi.org/10.1017/9781108768900.021 Published online by Cambridge University Press

Hence, one can thus relate the moments of $\rho_{\mathbf{E}}$ with the moments of $\rho_{\mathbf{C}}$ by taking $z \to \infty$ in Eq. (17.4), namely

$$\sum_{k=1}^{\infty} \frac{\tau(\mathbf{E}^k)}{z^k} = \sum_{k=1}^{\infty} \frac{\tau(\mathbf{C}^k)}{z^k} \left(1 + q \sum_{\ell=1}^{\infty} \tau(\mathbf{E}^\ell) z^{-\ell}\right)^k. \tag{17.10}$$

In particular, Eq. (17.10) yields the first three moments of $\rho_{\mathbf{E}}$:

$$\begin{aligned} \tau(\mathbf{E}) &= \tau(\mathbf{C}), \\ \tau(\mathbf{E}^2) &= \tau(\mathbf{C}^2) + q, \\ \tau(\mathbf{E}^3) &= \tau(\mathbf{C}^3) + 3q\tau(\mathbf{C}^2) + q^2. \end{aligned} \tag{17.11}$$

We thus see that the mean of $\mathbf{E}$ is equal to that of $\mathbf{C}$, whereas the variance of $\mathbf{E}$ is equal to that of $\mathbf{C}$ plus $q$. As expected, the spectrum of the sample covariance matrix $\mathbf{E}$ is always wider (for $q > 0$) than the spectrum of the population covariance matrix $\mathbf{C}$.

Another interesting expansion concerns the case where $q < 1$, such that $\mathbf{E}$ is invertible. Hence $\mathfrak{g}_{\mathbf{E}}(z)$ for $z \to 0$ is analytic and one readily finds

$$\mathfrak{g}_{\mathbf{E}}(z) \underset{z \to 0}{\longrightarrow} -\sum_{k=1}^{\infty} \tau\left(\mathbf{E}^{-k}\right) z^{k-1}. \tag{17.12}$$

This allows us to study the moments of $\mathbf{E}^{-1}$, which turn out to be important quantities for many applications. Using Eq. (17.5), we can actually relate the moments of the spectrum $\mathbf{E}^{-1}$ to those of $\mathbf{C}^{-1}$. Indeed, for $z \to 0$,

$$Z(z) = \frac{z}{1 - q - q \sum_{k=1}^{\infty} \tau\left(\mathbf{E}^{-k}\right) z^k}.$$

Hence, we obtain the following expansion:

$$\sum_{k=1}^{\infty} \tau\left(\mathbf{E}^{-k}\right) z^k = \sum_{k=1}^{\infty} \tau\left(\mathbf{C}^{-k}\right) \left(\frac{z}{1-q}\right)^k \left(\frac{1}{1 - \frac{q}{1-q} \sum_{\ell=1}^{\infty} \tau\left(\mathbf{E}^{-\ell}\right) z^\ell}\right)^k. \tag{17.13}$$

After a little work, we get

$$\tau\left(\mathbf{E}^{-1}\right) = \frac{\tau\left(\mathbf{C}^{-1}\right)}{1-q}, \qquad \tau\left(\mathbf{E}^{-2}\right) = \frac{\tau\left(\mathbf{C}^{-2}\right)}{(1-q)^2} + \frac{q\tau\left(\mathbf{C}^{-1}\right)^2}{(1-q)^3}. \tag{17.14}$$

We will discuss in Section 20.2.1 a direct application of these formulas: $\tau(\mathbf{E}^{-1})$ turns out to be related to the "out-of-sample" risk of an optimized portfolio of financial instruments.

**Exercise 17.1.1 The exponential moving average sample covariance matrix (EMA-SCM)**

Instead of measuring the sample covariance matrix using a flat average over a fixed time window $T$, one can compute the average using an exponential

https://doi.org/10.1017/9781108768900.021 Published online by Cambridge University Press

weighted moving average. Let us compute the spectrum of such a matrix in the null case of IID data. Imagine we have an infinite time series of vectors of size $N$ $\{\mathbf{x}_t\}$ for $t$ from minus infinity to now. We define the EMA-SCM (on time scale $\tau_c$) as

$$\mathbf{E}(t) = \gamma_c \sum_{t'=-\infty}^{t} (1-\gamma_c)^{t-t'} \mathbf{x}_{t'} \mathbf{x}_{t'}^T, \tag{17.15}$$

where $\gamma_c := 1/\tau_c$. Hence,

$$\mathbf{E}(t) = (1-\gamma_c)\mathbf{E}(t-1) + \gamma_c \mathbf{x}_{t'} \mathbf{x}_{t'}^T. \tag{17.16}$$

The second term on the right hand side can be thought of as a Wishart matrix with $T = 1$ (or $q = N$). Now, both $\mathbf{E}(t)$ and $\mathbf{E}(t-1)$ are equal in law so we write

$$\mathbf{E} \overset{\text{in law}}{=} (1-\gamma_c)\mathbf{E} + \gamma_c \mathbf{W}_{q=N}. \tag{17.17}$$

(a) Given that $\mathbf{E}$ and $\mathbf{W}$ are free, use the properties of the R-transform to get the equation

$$R_{\mathbf{E}}(x) = (1-\gamma_c) R_{\mathbf{E}}((1-\gamma_c)x) + \gamma_c(1 - N\gamma_c x). \tag{17.18}$$

(b) Take the limit $N \to \infty$, $\tau_c \to \infty$ with $q := N/\tau_c$ fixed to get the following differential equation for $R_{\mathbf{E}}(x)$:

$$R_{\mathbf{E}}(x) = -x \frac{\mathrm{d}}{\mathrm{d}x} R_{\mathbf{E}}(x) + \frac{1}{1-qx}. \tag{17.19}$$

(c) The definition of $\mathbf{E}$ is properly normalized, $\tau(\mathbf{E}) = 1$ [show this using Eq. (17.17)], so we have the initial condition $R(0) = 1$. Show that

$$R_{\mathbf{E}}(x) = -\frac{\log(1-qx)}{qx} \tag{17.20}$$

solves your equation with the correct initial condition. Compute the variance $\kappa_2(\mathbf{E})$.

(d) To compute the spectrum of eigenvalues of $\mathbf{E}$, one needs to solve a complex transcendental equation. First write $\mathfrak{z}(g)$, the inverse of $\mathfrak{g}(z)$. For $q = 1/2$ plot $\mathfrak{z}$ as a function of $g$ (for $-4 < g < 2$). You will see that there are values of $z$ that are never attained by $\mathfrak{z}(g)$, in other words $\mathfrak{g}(z)$ has no real solutions for these $z$. Numerically find complex solutions for $\mathfrak{g}(z)$ in that range. Plot the density of eigenvalues $\rho_{\mathbf{E}}(\lambda)$ given by Eq. (2.47). Plot also the density for a Wishart with the same mean and variance.

(e) Construct numerically the matrix $\mathbf{E}$ as in Eq. (17.15). Use $N = 1000$, $\tau_c = 2000$ and use at least 10 000 values for $t'$. Plot the eigenvalue distribution of your numerical $\mathbf{E}$ against the distribution found in (d).

https://doi.org/10.1017/9781108768900.021 Published online by Cambridge University Press

## 17.2 Temporal Correlations

### 17.2.1 General Case

A common problem in data analysis arises when samples are not independent. Intuitively, correlated samples are somehow redundant and the sample covariance matrix should behave as if we had observed not $T$ samples but an effective number $T^* < T$. Let us analyze more precisely the sample covariance matrix in the presence of correlated samples. We will start with the case when the true spatial correlations are zero, i.e. $\mathbf{C} = \mathbf{1}$. Our data can then be written in a rectangular $N \times T$ matrix $\mathbf{H}$ satisfying

$$\mathbb{E}[\mathbf{H}_{it}\mathbf{H}_{js}] = \delta_{ij}\mathbf{K}_{ts}, \tag{17.21}$$

where $\mathbf{K}$ is the $T \times T$ temporal covariance matrix that we assumed to be normalized as $\tau(\mathbf{K}) = 1$. Following the same arguments as in Section 4.2.4, we can write

$$\mathbf{H} = \mathbf{H}_0\mathbf{K}^{\frac{1}{2}}, \tag{17.22}$$

where $\mathbf{H}_0$ is a white rectangular matrix. So the sample covariance matrix becomes

$$\mathbf{E} = \frac{1}{T}\mathbf{H}\mathbf{H}^T = \frac{1}{T}\mathbf{H}_0\mathbf{K}\mathbf{H}_0^T. \tag{17.23}$$

Now this is not quite the free product of the matrix $\mathbf{K}$ and a white Wishart, but if we define the $(T \times T)$ matrix $\mathbf{F}$ as

$$\mathbf{F} = \frac{1}{N}\mathbf{H}^T\mathbf{H} = \frac{1}{N}\mathbf{K}^{\frac{1}{2}}\mathbf{H}_0^T\mathbf{H}_0\mathbf{K}^{\frac{1}{2}} \equiv \mathbf{K}^{\frac{1}{2}}\mathbf{W}_{1/q}\mathbf{K}^{\frac{1}{2}}, \tag{17.24}$$

then $\mathbf{F}$ is the free product of the matrix $\mathbf{K}$ and a white Wishart matrix with parameter $1/q$. Hence,

$$S_{\mathbf{F}}(t) = \frac{S_{\mathbf{K}}(t)}{1 + t/q}. \tag{17.25}$$

To find the S-transform of $\mathbf{E}$, we go back to Section 4.1.1, where we obtained Eq. (4.5) relating the Stieltjes transforms of $\mathbf{E}$ and $\mathbf{F}$. In terms of the T-transform, the relation is even simpler:

$$\mathfrak{t}_{\mathbf{F}}(z) = q\mathfrak{t}_{\mathbf{E}}(qz) \quad \Rightarrow \quad \zeta_{\mathbf{E}}(t) = q\zeta_{\mathbf{F}}(qt), \tag{17.26}$$

where the functions $\zeta(t)$ are the inverse T-transforms. Using the definition of the S-transform (Eq. (11.92)), we finally get

$$S_{\mathbf{E}}(t) = \frac{S_{\mathbf{K}}(qt)}{1 + qt}, \tag{17.27}$$

which can be expressed as a relation between inverse T-transforms:

$$\zeta_{\mathbf{E}}(t) = q(1 + t)\zeta_{\mathbf{K}}(qt). \tag{17.28}$$

https://doi.org/10.1017/9781108768900.021 Published online by Cambridge University Press

We can also write a subordination relation between the T-transforms:

$$q\mathfrak{t}_{\mathbf{E}}(z) = \mathfrak{t}_{\mathbf{K}}\left(\frac{z}{q(1+\mathfrak{t}_{\mathbf{E}}(z))}\right). \tag{17.29}$$

This is a general formula that we specialize to the case of exponential temporal correlations in the next section. Note that in the limit $z \to 0$, the above equation gives access to $\tau(\mathbf{E}^{-1})$. Using

$$\mathfrak{t}_{\mathbf{E}}(z) \underset{z\to 0}{=} -1 - \tau(\mathbf{E}^{-1})z + \mathcal{O}(z^2), \tag{17.30}$$

we find

$$\tau(\mathbf{E}^{-1}) = -\frac{1}{q\zeta_{\mathbf{K}}(-q)}. \tag{17.31}$$

### *17.2.2 Exponential Correlations*

The most common form of temporal correlation in experimental data is the decaying exponential, corresponding to a matrix $\mathbf{K}_{ts}$ in Eq. (17.21) given by

$$\mathbf{K}_{ts} := a^{|t-s|}, \tag{17.32}$$

where $1/\log(a)$ defines the temporal span of the correlations.

In Appendix A.3 we explicitly compute the S-transform of $\mathbf{K}$. The result reads

$$S_{\mathbf{K}}(t) = \frac{t+1}{\sqrt{1+(b^2-1)t^2} + bt}, \tag{17.33}$$

where $b := (1+a^2)/(1-a^2)$. From $S_{\mathbf{K}}$ one can also obtain $\zeta_{\mathbf{K}}$ and its inverse $\mathfrak{t}_{\mathbf{K}}$, which read

$$\zeta_{\mathbf{K}}(t) = \frac{\sqrt{1+(b^2-1)t^2}}{t} + b, \qquad \mathfrak{t}_{\mathbf{K}}(\zeta) = -\frac{1}{\sqrt{\zeta^2 - 2\zeta b + 1}}. \tag{17.34}$$

Combining Eq. (17.27) with Eq. (17.33), we get

$$S_{\mathbf{E}}(t) = \frac{1}{\sqrt{1+(b^2-1)(qt)^2} + bqt}. \tag{17.35}$$

From the S-transform, we find

$$\zeta_{\mathbf{E}}(t) = \frac{1+t}{tS_{\mathbf{E}}(t)} = \frac{1+t}{t}\left(\sqrt{1+(b^2-1)(qt)^2} + bqt\right), \tag{17.36}$$

which when inverted leads to a fourth order equation for $\mathfrak{t}_{\mathbf{E}}(z)$ that must be solved numerically, leading to the densities plotted in Fig. 17.1. However, one can obtain some information on $\tau(\mathbf{E}^{-1})$. From Eqs. (17.31) and (17.34), one obtains

$$\tau(\mathbf{E}^{-1}) = \frac{1}{\sqrt{q^2(b^2-1)+1} - bq} := \frac{1}{1-q^*}, \tag{17.37}$$

https://doi.org/10.1017/9781108768900.021 Published online by Cambridge University Press

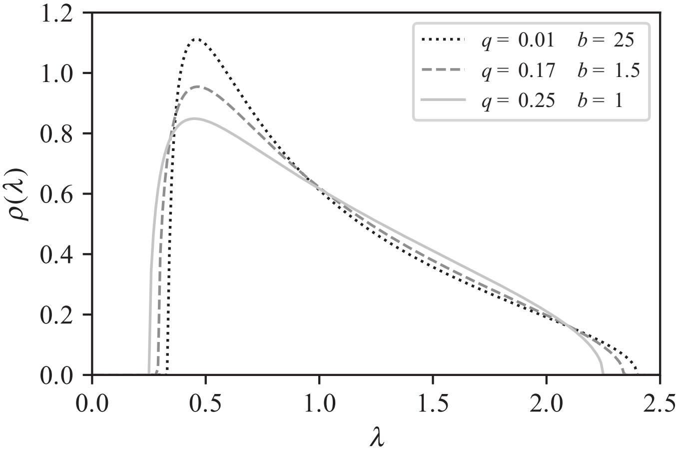


Figure 17.1 Density of eigenvalues for a sample covariance matrix with exponential temporal correlations for three choices of parameters $q$ and $b$ such that $qb = 0.25$. All three densities are normalized, have mean 1 and variance $\sigma_{\mathbf{E}}^2 = qb = 0.25$. The solid light gray one is the Marčenko–Pastur density ($q = 0.25$), the dotted black one is very close to the limiting density for $q \to 0$ with $\sigma^2 = bq$ fixed.

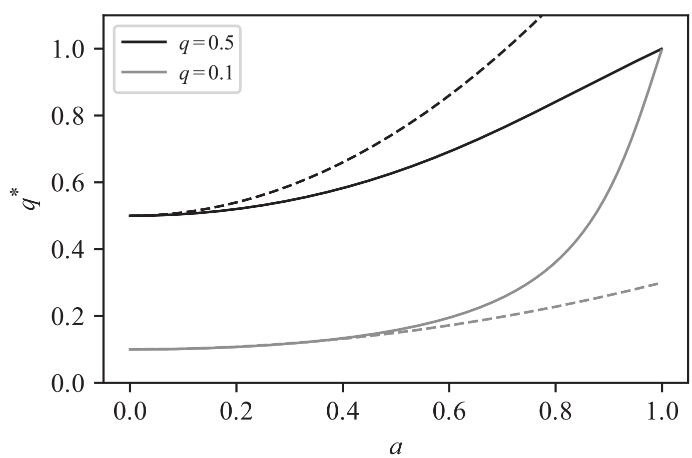


Figure 17.2 Effective value $q^*$ versus the one-lag autocorrelation coefficient $a$ for a sample covariance matrix with exponential temporal correlations shown for two values of $q$. The dashed lines indicate the approximation (valid at small $a$) $q^* = q(1 + 2a^2)$. The approximation means that, for 10% autocorrelation, $q^*$ is only 2% greater than $q$.

where $q^* = N/T^*$ defines the effective length of the time series, reduced by temporal correlations (compare with Eq. (17.14) with $\mathbf{C} = \mathbf{1}$). Figure 17.2 shows $q^*$ as a function of $a$. As expected, $q^* = q$ for $a = 0$ (no temporal correlations), whereas $q^* \to 1$ when $a \to 1$, i.e. when $\tau_c \to \infty$. In this limit, $\mathbf{E}$ becomes singular.

https://doi.org/10.1017/9781108768900.021 Published online by Cambridge University Press

Looking at Eq. (17.35), one notices that when $b \gg 1$ (corresponding to $a \to 1$, i.e. slowly decaying correlations), the S-transform depends on $b$ and $q$ only through the combination $qb$. One can thus define a new limiting distribution corresponding to the limit $q \to 0$, $b \to \infty$ with $qb = \sigma^2$ (which turns out to be the variance of the distribution, see below). The S-transform of this limiting distribution is given by

$$S(t) = \frac{1}{\sqrt{1 + (\sigma^2 t)^2} + \sigma^2 t}, \tag{17.38}$$

while the equation for the T-transform boils down to a cubic equation that reads:

$$z^2 \mathfrak{t}^2(z) - 2\sigma^2 z \mathfrak{t}^2(z)(1 + \mathfrak{t}(z)) = (1 + \mathfrak{t}(z))^2. \tag{17.39}$$

The corresponding R-transform is

$$\begin{aligned} R(z) &= \frac{1}{\sqrt{1 - 2\sigma^2 z}} \\ &= 1 + \sigma^2 z + \frac{3}{2}\sigma^4 z^2 + O(z^3). \end{aligned} \tag{17.40}$$

The last equation gives its first three cumulants: its average is equal to one, its variance is $\sigma^2$ as announced above, and its skewness is $\kappa_3 = \frac{3}{2}\sigma^4$. We notice that this skewness is larger than that of a white Wishart with the same variance ($q = \sigma^2$) for which $\kappa_3 = \sigma^4$. The equations for the Stieltjes $\mathfrak{g}(z)$ and the T-transform are both cubic equations. The corresponding distribution of eigenvalues is shown in Figure 17.3. Note that, unlike the Marčenko–Pastur, there is always a strictly positive lower edge of the spectrum $\lambda_- > 0$ and no Dirac at zero even when $\sigma^2 > 1$. Unfortunately, the equation giving $\lambda_\pm$ is a fourth order equation that does not have a concise solution.

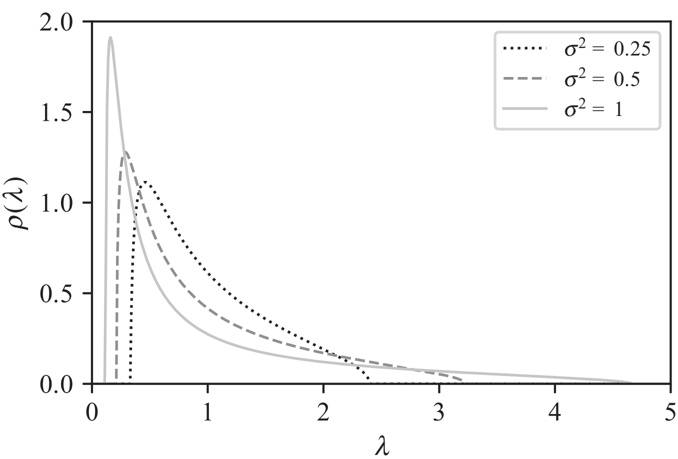


Figure 17.3 Density of eigenvalues for the limiting distribution of sample covariance matrix with exponential temporal correlations $\mathbf{W}_{\sigma^2}$ for three choices of the parameter $\sigma^2$: 0.25, 0.5 and 1.

https://doi.org/10.1017/9781108768900.021 Published online by Cambridge University Press

An intuitive way to understand this particular random matrix ensemble is to consider $N$ independent Ornstein–Uhlenbeck processes with the same correlation time $\tau_c$ that we record over a long time $T$. We sample the data at interval $\Delta$, such that the total number of observations is $T/\Delta$. We then construct a sample covariance matrix of the $N$ variables from these observations. If $\Delta \gg \tau_c$, then each sample can be considered independent and the sample covariance matrix will be a Marčenko–Pastur with $q = N\Delta/T$. But if we "oversample" at intervals $\Delta \ll \tau_c$, such that our observations are strongly correlated, then the resulting sample covariance matrix no longer depends on $\Delta$ but only on $\tau_c$. The sample covariance matrix converges in this case to our new random matrix characterized by Eq. (17.38), with parameter $\sigma^2 = qb = N\tau_c/T$.

### 17.2.3 Spatial and Temporal Correlations

In the general case where spatial and temporal correlations exist, the sample covariance matrix can be written as

$$\mathbf{E} = \frac{1}{T}\mathbf{H}\mathbf{H}^T = \frac{1}{T}\mathbf{C}^{\frac{1}{2}}\mathbf{H}_0\mathbf{K}\mathbf{H}_0^T\mathbf{C}^{\frac{1}{2}}, \tag{17.41}$$

using the same notations as above. After similar manipulations, the S-transform of $\mathbf{E}$ is found to be given by

$$S_{\mathbf{E}}(t) = \frac{S_{\mathbf{C}}(t)S_{\mathbf{K}}(qt)}{1+qt}, \tag{17.42}$$

which leads to

$$\zeta_{\mathbf{E}}(t) = qt\zeta_{\mathbf{C}}(t)\zeta_{\mathbf{K}}(qt), \tag{17.43}$$

or, in terms of T-transforms,

$$q\mathrm{t}_{\mathbf{E}}(z) = \mathrm{t}_{\mathbf{K}}\left(\frac{z}{q\mathrm{t}_{\mathbf{E}}(z)\zeta_{\mathbf{C}}(\mathrm{t}_{\mathbf{E}}(z))}\right). \tag{17.44}$$

When $\mathbf{C} = \mathbf{1}$, $\zeta_{\mathbf{C}}(t) = (1+t)/t$ and one recovers Eq. (17.29). Specializing to the case of exponential correlations in the limit $q \to 0$, $a \to 1$, $qb = \sigma^2$, we obtain the following equation for the T-transform of the limiting distribution, now for an arbitrary covariance matrix $\mathbf{C}$:

$$z^2 - 2\sigma^2 z\mathrm{t}_{\mathbf{E}}(z)\zeta_{\mathbf{C}}(\mathrm{t}_{\mathbf{E}}(z)) = \zeta^2_{\mathbf{C}}(\mathrm{t}_{\mathbf{E}}(z)), \tag{17.45}$$

where we used $\mathrm{t}_{\mathbf{K}}(z) = -1/\sqrt{z^2 - 2zb + 1}$. When $\mathbf{C} = \mathbf{1}$, one recovers Eq. (17.39). When $\mathbf{C}$ is an inverse-Wishart matrix, $\zeta_{\mathbf{C}}(t) = (t+1)/t(1-pt)$, the equation for $\mathrm{t}_{\mathbf{E}}(z)$ is of fourth order.

Note finally that Eq. (17.44), in the limit $z \to 0$, yields a simple generalization of Eq. (17.31) that reads

$$\tau(\mathbf{E}^{-1}) = -\frac{\tau(\mathbf{C}^{-1})}{q\zeta_{\mathbf{K}}(-q)}. \tag{17.46}$$

https://doi.org/10.1017/9781108768900.021 Published online by Cambridge University Press

Comparing with Eq. (17.14) allows us to define an effective length of the time series which, interestingly, is *independent* of **C** and reads

$$q^* := \frac{N}{T^*} = 1 + q\zeta_{\mathbf{K}}(-q). \tag{17.47}$$

**Exercise 17.2.1 On the futility of oversampling**

Consider data consisting of $N$ variables (columns) with true correlation $\mathbf{C}$ and $T$ independent observations (rows). Instead of computing the sample covariance matrix with these $T$ observations, we repeat each one $m$ times and sum over $mT$ columns. Obviously the redundant columns should not change the sample covariance matrix, hence it should have the same spectrum as the one using only the original $T$ observations.

(a) The redundancy of columns can be modeled as a temporal correlation with an $mT \times mT$ covariance matrix $\mathbf{K}$ that is block diagonal with $T$ blocks of size $K$ and all the values within one block equal to 1 and zero outside the blocks. Show that this matrix has $T$ eigenvalues equal to $m$ and $(T-1)m$ zero eigenvalues.

(b) Compute $\mathfrak{t}_{\mathbf{K}}(z)$ for this model.

(c) Show that $S_{\mathbf{K}}(t) = (1+t)/(1+mt)$.

(d) If we include the redundant columns we have a value of $q_m = N/(mT)$, but we need to take temporal correlations into account so $S_{\mathbf{E}}(t) = S_{\mathbf{C}}(t)S_{\mathbf{K}}(q_m t)/(1+q_m t)$. Show that in this case $S_{\mathbf{E}}(t) = S_{\mathbf{C}}(t)/(1+qt)$ with $q = N/T$, which is the result without the redundant columns.

## 17.3 Time Dependent Variance

Another common and important situation is when the $N$ correlated time series are heteroskedastic, i.e. have a time dependent variance. More precisely, we consider a model where

$$x_i^t = \sigma_t \mathbf{H}_{it}, \tag{17.48}$$

where $\sigma_t$ is time dependent, and

$$\mathbb{E}[\mathbf{H}_{it}\mathbf{H}_{js}] = \delta_{ts}\mathbf{C}_{ij}, \tag{17.49}$$

i.e. $x_i^t$ is the product of a time dependent factor $\sigma_t$ and a random variable with a general correlation structure $\mathbf{C}$ but no time correlations. The SCM $\mathbf{E}$ can be expressed as

$$\mathbf{E} = \sum_{t=1}^{T} \mathbf{P}_t, \qquad \mathbf{P}_t := \frac{1}{T}\sigma_t^2 \mathbf{H}_t \mathbf{H}_t^T, \tag{17.50}$$

https://doi.org/10.1017/9781108768900.021 Published online by Cambridge University Press

where each $\mathbf{P}_t$ is a rank-1 matrix with a non-zero eigenvalue that converges, when $N$ and $T$ tend to infinity, to $q\sigma_t^2\tau(\mathbf{C})$ with, as always, $q = N/T$.

We will first consider the case $\mathbf{C} = \mathbf{1}$, i.e. a structureless covariance matrix. In this case, the vectors $\mathbf{x}^t$ are rotationally invariant, the matrix $\mathbf{E}$ can be viewed as the free sum of a large number of rank-1 matrices, each with a non-zero eigenvalue equal to $q\sigma_t^2$. Hence,

$$R_{\mathbf{E}}(g) = \sum_{t=1}^{T} R_t(g). \tag{17.51}$$

To compute the R-transform of the matrix $\mathbf{E}$ we need to compute the R-transform of a rank-1 matrix. Note that since there are $T$ terms in the sum, we will need to know $R_t(g)$ including correction of order $1/N$:

$$\mathfrak{g}_t(z) = \frac{1}{N}\left(\frac{N-1}{z} + \frac{1}{z - q\sigma_t^2}\right) = \frac{1}{z} + \frac{1}{N}\frac{q\sigma_t^2}{z(z - q\sigma_t^2)}. \tag{17.52}$$

Inverting to first order in $1/N$ we find

$$\mathfrak{z}_t(g) = \frac{1}{g} + \frac{1}{N}\frac{q\sigma_t^2}{1 - q\sigma_t^2 g}. \tag{17.53}$$

Now, since $R(z) = \mathfrak{z}(g) - 1/z$, we find

$$R_{\mathbf{E}}(g) = \frac{1}{T}\sum_{t=1}^{T} \frac{\sigma_t^2}{1 - q\sigma_t^2 g}. \tag{17.54}$$

The fluctuations of $\sigma_t^2$ can be stochastic or deterministic. In the large $T$ limit we can encode them with a probability density $P(s)$ for $s = \sigma^2$ and convert the sum into an integral, leading to[1]

$$R_{\mathbf{E}}(g) = \int_0^\infty \frac{sP(s)}{1 - qsg}\mathrm{d}s. \tag{17.55}$$

Note that if the variance is always 1 (i.e. $P(s) = \delta(s-1)$), we recover the R-transform of a Wishart matrix of parameter $q$:

$$R_q(g) = \frac{1}{1 - qg}. \tag{17.56}$$

In the general case, the R-transform of $\mathbf{E}$ is simply related to the T-transform of the distribution of $s$:

$$R_{\mathbf{E}}(g) = \mathfrak{t}_s\left(\frac{1}{qg}\right). \tag{17.57}$$

[1] When the distribution of $s$ is bounded, the integral (17.55) always converges for small enough $g$ and the R-transform is well defined near zero. For unbounded $s$, the R-transform can be singular at zero indicating that the distribution of eigenvalues doesn't have an upper edge.

In the more general case where $\mathbf{C}$ is not the identity matrix, one can again write the SCM as $\widetilde{\mathbf{E}} = \mathbf{C}^{\frac{1}{2}}\mathbf{E}\mathbf{C}^{\frac{1}{2}}$, where $\mathbf{E}$ corresponds to the case $\mathbf{C} = \mathbf{1}$ that we just treated. Hence, using the fact that $\mathbf{C}$ and $\mathbf{E}$ are mutually free, the S-transform of $\widetilde{\mathbf{E}}$ is simply given by

$$S_{\widetilde{\mathbf{E}}}(t) = S_{\mathbf{C}}(t)S_{\mathbf{E}}(t). \tag{17.58}$$

Another way to treat the problem is to view the fluctuating variance as a diagonal temporal covariance matrix with entries drawn from $P(s)$. Following Section 17.2.3, we can write

$$S_{\mathbf{E}}(t) = \frac{S_s(qt)}{1+qt}, \quad S_{\widetilde{\mathbf{E}}}(t) = \frac{S_{\mathbf{C}}(t)S_s(qt)}{1+qt}, \tag{17.59}$$

with $S_s(t)$ the S-transform associated with $\mathrm{t}_s(\zeta)$.

A particular case of interest for financial applications is when $P(s)$ is an inverse-gamma distribution. When $\mathbf{x}^t$ is a Gaussian multivariate vector, one obtains for $\sigma_t\mathbf{x}^t$ a *Student multivariate distribution* (see bibliographical notes for more on this topic).

## 17.4 Empirical Cross-Covariance Matrices

Let us now consider two time series $\mathbf{x}^t$ and $\mathbf{y}^t$, each of length $T$, but of different dimensions, respectively $N_1$ and $N_2$. The empirical cross-covariance matrix is an $N_1 \times N_2$ rectangular matrix defined as

$$\mathbf{E}_{xy} = \frac{1}{T}\sum_{t=1}^{T} \mathbf{x}^t(\mathbf{y}^t)^T. \tag{17.60}$$

Let us assume that the "true" cross-covariance matrix $\mathbb{E}[\mathbf{x}\mathbf{y}^T]$ is zero, i.e. that there are no true cross-correlations between our two sets of variables. What is the singular value spectrum of $\mathbf{E}_{xy}$ in this case?

As with SCM that are described by the Marčenko–Pastur law when $N, T \to \infty$ with a fixed ratio $q = N/T$, we expect that some non-trivial results will appear in the limit $N_1, N_2, T \to \infty$ with $q_1 = N_1/T$ and $q_2 = N_2/T$ finite. A convenient way to perform this analysis is to consider the eigenvalues of the $N_1 \times N_1$ matrix $\mathbf{M}_{xy} = \mathbf{E}_{xy}\mathbf{E}_{xy}^T$, which are equal to the square of the singular values $s$ of $\mathbf{E}_{xy}$.

The matrix $\mathbf{M}_{xy}$ shares the same non-zero eigenvalues as those of $\widehat{\mathbf{E}}_x\widehat{\mathbf{E}}_y$, where $\widehat{\mathbf{E}}_x$ and $\widehat{\mathbf{E}}_y$ are the dual $T \times T$ sample covariance matrices:

$$\widehat{\mathbf{E}}_x = \mathbf{x}^T\mathbf{x}, \qquad \widehat{\mathbf{E}}_y = \mathbf{y}^T\mathbf{y}. \tag{17.61}$$

Hence one can compute the spectral density of $\mathbf{M}_{xy}$ using the free product formalism and infer the spectrum of the product $\widehat{\mathbf{E}}_x\widehat{\mathbf{E}}_y$. However, the result will depend on the "true" covariance matrices of $\mathbf{x}$ and $\mathbf{y}$, which are usually unknown in practical applications.

A way to obtain a universal result is to consider the sample-normalized principal components of $\mathbf{x}$ and of $\mathbf{y}$, which we call $\widetilde{\mathbf{x}}$ and $\widetilde{\mathbf{y}}$, such that the corresponding dual covariance matrix $\widehat{\mathbf{E}}_{\widetilde{x}}$ has $N_1$ eigenvalues exactly equal to 1 and $T - N_1$ eigenvalues exactly equal to zero, whereas $\widehat{\mathbf{E}}_{\widetilde{y}}$ has $N_2$ eigenvalues exactly equal to 1 and $T - N_2$ eigenvalues exactly equal to zero. This is precisely the problem studied in Section 15.4.2. The singular value spectrum of $\mathbf{E}_{xy}$ is thus given by

$$\rho(s) = \max(q_1 + q_2 - 1, 0)\delta(s-1) + \mathrm{Re}\,\frac{\sqrt{(s^2-\gamma_-)(\gamma_+ - s^2)}}{\pi s(1-s^2)}, \tag{17.62}$$

https://doi.org/10.1017/9781108768900.021 Published online by Cambridge University Press

where $\gamma_\pm$ are given by

$$\gamma_\pm = q_1 + q_2 - 2q_1q_2 \pm 2\sqrt{q_1q_2(1-q_1)(1-q_2)}, \quad 0 \le \gamma_\pm \le 1. \tag{17.63}$$

The allowed $s$'s are all between 0 and 1, as they should be, since these singular values can be interpreted as correlation coefficients between some linear combination of the $\mathbf{x}$'s and some other linear combination of the $\mathbf{y}$'s.

In the limit $T \to \infty$ at fixed $N_1$, $N_2$, all singular values collapse to zero, as they should since there are no true correlations between $\mathbf{x}$ and $\mathbf{y}$. The allowed band in the limit $q_1, q_2 \to 0$ becomes

$$s \in \left[|\sqrt{q_1} - \sqrt{q_2}|, \sqrt{q_1} + \sqrt{q_2}\right],$$

showing that for fixed $N_1, N_2$, the order of magnitude of allowed singular values decays as $T^{-\frac{1}{2}}$. The above result allows one to devise precise statistical tests to detect "true" cross-correlations between sets of variables.

### *Bibliographical Notes*

- The subject of this chapter is treated in several books, see e.g.
  - Z. Bai and J. W. Silverstein. *Spectral Analysis of Large Dimensional Random Matrices*. Springer-Verlag, New York, 2010,
  - A. M. Tulino and S. Verdú. *Random Matrix Theory and Wireless Communications*. Now publishers, Hanover, Mass., 2004,
  - L. Pastur and M. Scherbina. *Eigenvalue Distribution of Large Random Matrices*. American Mathematical Society, Providence, Rhode Island, 2010,
  - R. Couillet and M. Debbah. *Random Matrix Methods for Wireless Communications*. Cambridge University Press, Cambridge, 2011.
- The initial "historical" paper is of course
  - V. A. Marchenko and L. A. Pastur. Distribution of eigenvalues for some sets of random matrices. *Matematicheskii Sbornik*, 114(4):507–536, 1967,

  with many posterior rediscoveries – as for example in[2]
  - A. Crisanti and H. Sompolinsky. Dynamics of spin systems with randomly asymmetric bonds: Langevin dynamics and a spherical model. *Physical Review A*, 36:4922–4939, 1987,
  - A. Sengupta and P. P. Mitra. Distributions of singular values for some random matrices. *Physical Review E*, 60(3):3389, 1999.
- The universality of Eq. (17.5) is discussed in
  - J. W. Silverstein and Z. Bai. On the empirical distribution of eigenvalues of a class of large dimensional random matrices. *Journal of Multivariate Analysis*, 54(2):175–192, 1995.
- The case of power-law distributed variables and deviations from Marčenko–Pastur are discussed in

[2] In fact, Crisanti and Sompolinsky [1987] themselves cite an unpublished work of D. Movshovitz and H. Sompolinsky.

- S. Belinschi, A. Dembo, and A. Guionnet. Spectral measure of heavy tailed band and covariance random matrices. *Communications in Mathematical Physics*, 289(3):1023–1055, 2009,
- G. Biroli, J.-P. Bouchaud, and M. Potters. On the top eigenvalue of heavy-tailed random matrices. *Europhysics Letters (EPL)*, 78(1):10001, 2007,
- A. Auffinger, G. Ben Arous, and S. Péché. Poisson convergence for the largest eigenvalues of heavy tailed random matrices. *Annales de l'I.H.P. Probabilités et statistiques*, 45(3):589–610, 2009.


- The case with spatial and temporal correlations was studied in


  - A. Sengupta and P. P. Mitra. Distributions of singular values for some random matrices. *Physical Review E*, 60(3):3389, 1999,
  - Z. Burda, J. Jurkiewicz, and B. Wacław. Spectral moments of correlated Wishart matrices. *Physical Review E*, 71:026111, 2005.


- Multivariate Student (or elliptical) variables and their associated sample covariance matrices are studied in


  - G. Biroli, J.-P. Bouchaud, and M. Potters. The Student ensemble of correlation matrices: eigenvalue spectrum and Kullback-Leibler entropy. *preprint arXiv:0710.0802*, 2007,
  - N. El Karoui et al. Concentration of measure and spectra of random matrices: Applications to correlation matrices, elliptical distributions and beyond. *The Annals of Applied Probability*, 19(6):2362–2405, 2009.


- Large cross-correlation matrices and corresponding null-hypothesis statistical tests are studied in


  - I. M. Johnstone. Multivariate analysis and Jacobi ensembles: Largest eigenvalue, TracyWidom limits and rates of convergence. *The Annals of Statistics*, 36(6):2638–2716, 2008,
  - J.-P. Bouchaud, L. Laloux, M. A. Miceli, and M. Potters. Large dimension forecasting models and random singular value spectra. *The European Physical Journal B*, 55(2):201–207, 2007,
  - Y. Yang and G. Pan. Independence test for high dimensional data based on regularized canonical correlation coefficients. *The Annals of Statistics*, 43(2):467–500, 2015.


- For an early derivation of Eq. (17.62) without the use of free probability methods, see


  - K. W. Wachter. The limiting empirical measure of multiple discriminant ratios. *The Annals of Statistics*, 8(5):937–957, 1980.