Random Matrix Theory and Correlation Estimation

Jim Gatheral

Baruch College
The City University of New York

Baruch College Mathematics Society,
February 24, 2015

Motivation

We would like to understand

- what is random matrix theory. (RMT)

Motivation

We would like to understand

- what is random matrix theory. (RMT)
- how to apply RMT to the estimation of covariance matrices.

Motivation

We would like to understand

- what is random matrix theory. (RMT)
- how to apply RMT to the estimation of covariance matrices.
- whether the resulting covariance matrix performs better than (for example) the Barra covariance matrix.

Outline

Outline

1. Random matrix theory
   - Random matrix examples
   - Wigner's semicircle law
   - The Marčenko-Pastur density
   - The Tracy-Widom law
   - Impact of fat tails
2. Estimating correlations
   - Uncertainty in correlation estimates.
   - Example with SPX stocks
   - A recipe for filtering the sample correlation matrix

Outline

1. Random matrix theory
   - Random matrix examples
   - Wigner's semicircle law
   - The Marčenko-Pastur density
   - The Tracy-Widom law
   - Impact of fat tails
2. Estimating correlations
   - Uncertainty in correlation estimates.
   - Example with SPX stocks
   - A recipe for filtering the sample correlation matrix
3. Comparison with Barra
   - Comparison of eigenvectors
   - The minimum variance portfolio
     - Comparison of weights
     - In-sample and out-of-sample performance

Outline

1. Random matrix theory
    - Random matrix examples
    - Wigner's semicircle law
    - The Marčenko-Pastur density
    - The Tracy-Widom law
    - Impact of fat tails
2. Estimating correlations
    - Uncertainty in correlation estimates.
    - Example with SPX stocks
    - A recipe for filtering the sample correlation matrix
3. Comparison with Barra
    - Comparison of eigenvectors
    - The minimum variance portfolio
        - Comparison of weights
        - In-sample and out-of-sample performance
4. Conclusions

Outline

Example 1: Normal random symmetric matrix

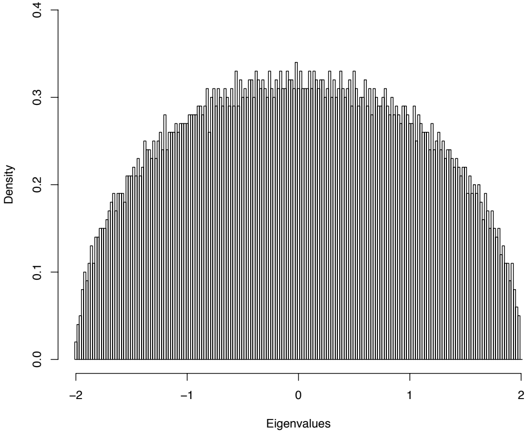
- Generate a 5,000 x 5,000 random symmetric matrix with entries $a_{ij} \sim N(0,1)$.
- Compute eigenvalues.
- Draw a histogram.

Here's some R-code to generate a symmetric random matrix whose off-diagonal elements have variance $1/N$:

```
n <- 5000;
m <- array(rnorm(n^2),c(n,n));
m2 <- (m+t(m))/sqrt(2*n);# Make m symmetric
lambda <- eigen(m2, symmetric=T, only.values = T);
e <- lambda$values;
hist(e,breaks=seq(-2.01,2.01,.02),
    main=NA, xlab="Eigenvalues",freq=F)
```

Example 1: continued

Here's the result:



Example 2: Uniform random symmetric matrix

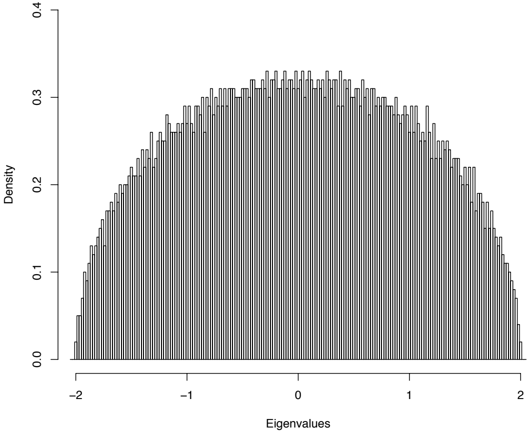
- Generate a 5,000 x 5,000 random symmetric matrix with entries $a_{ij} \sim \mathit{Uniform}(0,1)$.
- Compute eigenvalues.
- Draw a histogram.

Here's some R-code again:

```
n <- 5000;
mu <- array(runif(n^2),c(n,n));
mu2 <-sqrt(12)*(mu+t(mu)-1)/sqrt(2*n);
lambdau <- eigen(mu2, symmetric=T, only.values = T);
eu <- lambdau$values;
hist(eu,breaks=seq(-2.05,2.05,.02),main=NA,xlab="Eigenvalue
eu <- lambdau$values;
histeu<-hist(eu,breaks=seq(-2.01,2.01,0.02),
    main=NA, xlab="Eigenvalues",freq=F)
```

Example 2: continued

Here's the result:



What do we see?

We note a striking pattern: the density of eigenvalues is a semicircle!

Wigner's semicircle law

Consider an $N \times N$ matrix $\tilde{\mathbf{A}}$ with entries $\tilde{a}_{ij} \sim N(0, \sigma^2)$. Define

$$\mathbf{A}_N = \frac{1}{\sqrt{2\,N}} \left\{ \tilde{\mathbf{A}} + \tilde{\mathbf{A}}' \right\}$$

Then $\mathbf{A}_N$ is symmetric with

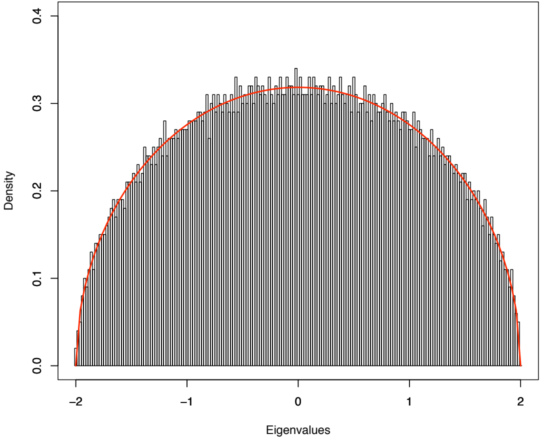
$$\mathrm{Var}[a_{ij}] = \begin{cases} \sigma^2/N & \text{if } i \neq j \\ 2\,\sigma^2/N & \text{if } i = j \end{cases}$$

The density of eigenvalues of $\mathbf{A}_N$ is given by

$$\begin{aligned} \rho_N(\lambda) \quad &:= \quad \frac{1}{N} \sum_{i=1}^{N} \delta(\lambda - \lambda_i) \\ &\xrightarrow[N \to \infty]{} \begin{cases} \frac{1}{2\,\pi\,\sigma^2} \sqrt{4\,\sigma^2 - \lambda^2} & \text{if } |\lambda| \leq 2\,\sigma \\ 0 & \text{otherwise.} \end{cases} \quad =: \rho(\lambda) \end{aligned}$$

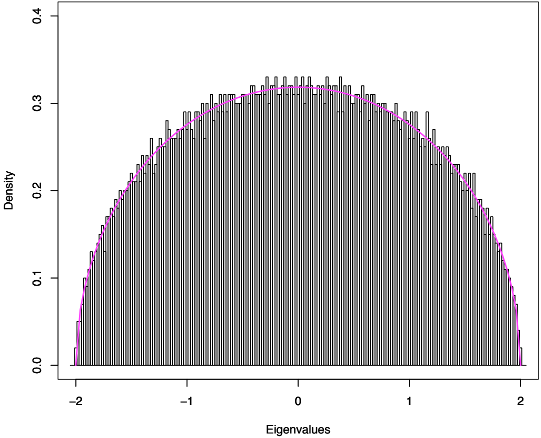
Example 1: Normal random matrix with Wigner density

Now superimpose the Wigner semicircle density:



Example 2: Uniform random matrix with Wigner density

Again superimpose the Wigner semicircle density:



Random correlation matrices

Suppose we have M stock return series with T elements each. The elements of the $M \times M$ empirical correlation matrix E are given by

$$E_{ij} = \frac{1}{T} \sum_{t}^{T} x_{it}\, x_{jt}$$

where x_{it} denotes the tth return of stock i, normalized by standard deviation so that $\mathrm{Var}[x_{it}] = 1$.

In matrix form, this may be written as

$$\mathbf{E} = \mathbf{H}\,\mathbf{H}'$$

where $\mathbf{H}$ is the $M \times T$ matrix whose rows are the time series of returns, one for each stock.

Eigenvalue spectrum of random correlation matrix

Suppose the entries of **H** are random with variance σ^2. Then, in the limit $T, M \to \infty$ keeping the ratio $Q := T/M \geq 1$ constant, the density of eigenvalues of **E** is given by

$$\rho(\lambda) = \frac{Q}{2\,\pi\,\sigma^2} \frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{\lambda}$$

where the maximum and minimum eigenvalues are given by

$$\lambda_\pm = \sigma \left(1 \pm \sqrt{\frac{1}{Q}}\right)^2 .$$

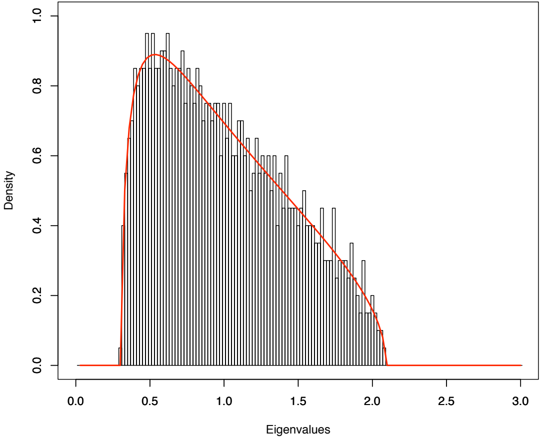
$\rho(\lambda)$ is known as the Marčenko-Pastur density.

Example: IID random normal returns

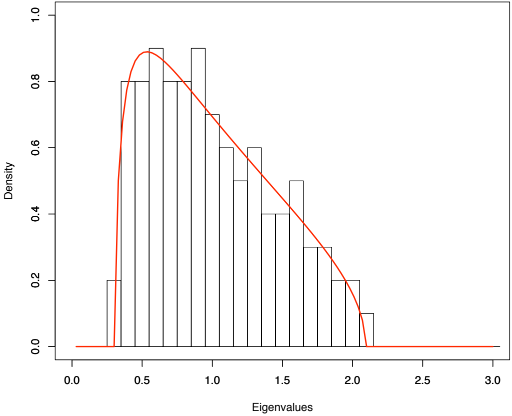
Here's some R-code again:

```
t <- 5000;
m <- 1000;
h <- array(rnorm(m*t),c(m,t)); # Time series in rows
e <- h %*% t(h)/t; # Form the correlation matrix
lambdae <- eigen(e, symmetric=T, only.values = T);
ee <- lambdae$values;
hist(ee,breaks=seq(0.01,3.01,.02),
main=NA,xlab="Eigenvalues",freq=F)
```

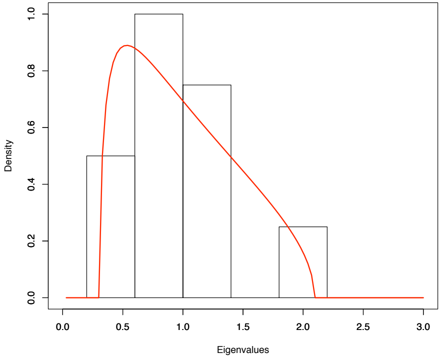
Here's the result with the Marčenko-Pastur density superimposed:



Here's the result with $M = 100$, $T = 500$ (again with the Marčenko-Pastur density superimposed):



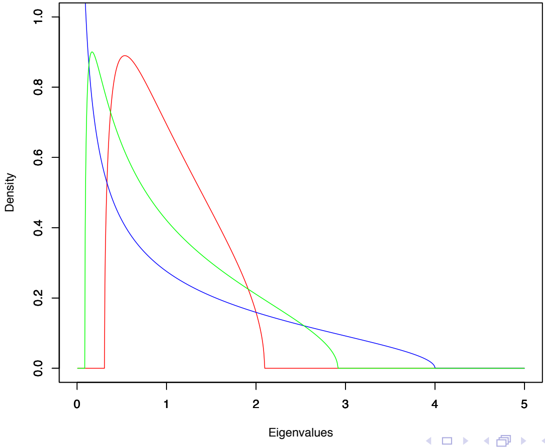
...and again with $M = 10$, $T = 50$:



We see that even for rather small matrices, the theoretical limiting density approximates the actual density very well.

Some Marčenko-Pastur densities

The Marčenko-Pastur density depends on $Q = T/M$. Here are graphs of the density for $Q = 1$ (blue), 2 (green) and 5 (red).



Distribution of the largest eigenvalue

- For applications where we would like to know where the random bulk of eigenvalues ends and the spectrum of eigenvalues corresponding to true information begins, we need to know the distribution of the largest eigenvalue.
- The distribution of the largest eigenvalue of a random correlation matrix is given by the Tracy-Widom law.

$$\Pr(T\,\lambda_{max} < \mu_{TM} + s\,\sigma_{TM}) = F_1(s)$$

with

$$\begin{aligned}
\mu_{TM} &= \left(\sqrt{T - 1/2} + \sqrt{M - 1/2}\right)^2 \\
\sigma_{TM} &= \left(\sqrt{T - 1/2} + \sqrt{M - 1/2}\right)\left(\frac{1}{\sqrt{T - 1/2}} + \frac{1}{\sqrt{M - 1/2}}\right)^{1/3}
\end{aligned}$$

Fat-tailed random matrices

- So far, we have considered matrices whose entries are either Gaussian or drawn from distributions with finite moments.
- Suppose that entries are drawn from a fat-tailed distribution such as Lévy-stable.
  - This is of practical interest because we know that stock returns follow a cubic law and so are fat-tailed.
- Bouchaud et. al. find that fat tails can massively increase the maximum eigenvalue in the theoretical limiting spectrum of the random matrix.
  - Where the distribution of matrix entries is extremely fat-tailed (Cauchy for example) , the semi-circle law no longer holds.

Sampling error

- Suppose we compute the sample correlation matrix of M stocks with T returns in each time series.
- Further suppose that the true correlation matrix were the identity matrix. What would we expect the greatest sample correlation to be?
- For $N(0,1)$ distributed returns, the typical maximum correlation ρ_{max} should satisfy:

$$\frac{2}{M\left(M-1\right)} \sim N\left(-\rho_{max}\sqrt{T}\right)$$

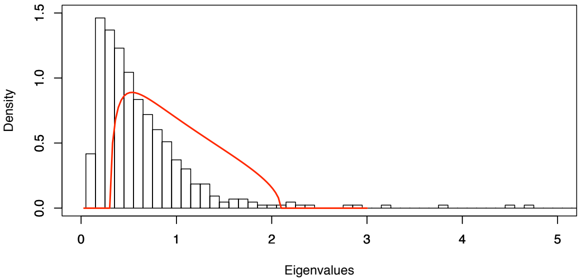
  - With $M = 500, T = 1000$, we obtain $\rho_{max} \approx 0.14$.
- So, sampling error induces spurious (and potentially significant) correlations between stocks!

An experiment with real data

- We take 431 stocks in the SPX index for which we have $2,155 = 5 \times 431$ consecutive daily returns.
  - Thus, in this case, $M = 431$ and $T = 2,155$. $Q = T/M = 5$.
  - There are $M(M-1)/2 = 92,665$ distinct entries in the correlation matrix to be estimated from $2,155 \times 431 = 928,805$ data points.
  - With these parameters, we would expect the maximum error in our correlation estimates to be around 0.09.
- First, we compute the eigenvalue spectrum and superimpose the Marčenko Pastur density with $Q = 5$.

The eigenvalue spectrum of the sample correlation matrix

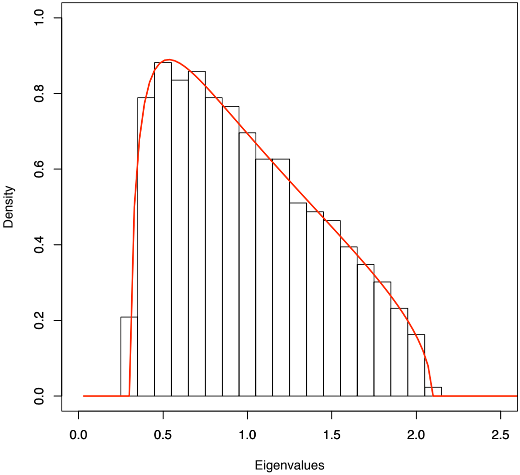
Here's the result:



Note that the top eigenvalue is 105.37 – way off the end of the chart! The next biggest eigenvalue is 18.73.

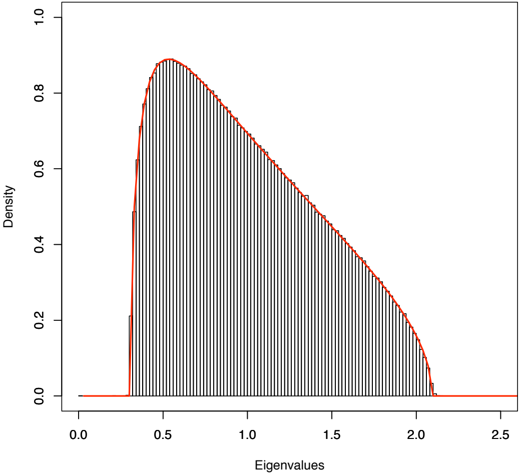
With randomized return data

Suppose we now shuffle the returns in each time series. We obtain:



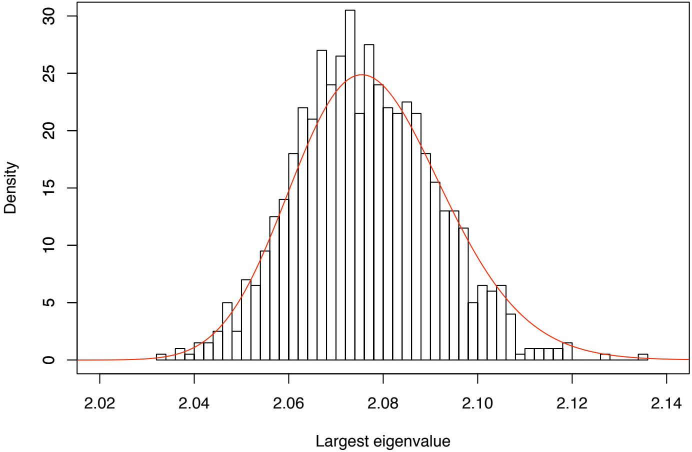
Repeat 1,000 times and average

Repeating this 1,000 times gives:



Distribution of largest eigenvalue

We can compare the empirical distribution of the largest eigenvalue with the Tracy-Widom density (in red):



Interim conclusions

From this simple experiment, we note that:

- Even though return series are fat-tailed,
    - the Marčenko-Pastur density is a very good approximation to the density of eigenvalues of the correlation matrix of the randomized returns.
    - the Tracy-Widom density is a good approximation to the density of the largest eigenvalue of the correlation matrix of the randomized returns.
- The Marčenko-Pastur density does not remotely fit the eigenvalue spectrum of the sample correlation matrix from which we conclude that there is nonrandom structure in the return data.
- We may compute the theoretical spectrum arbitrarily accurately by performing numerical simulations.

Problem formulation

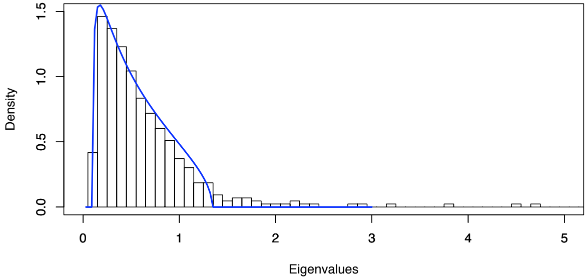
Which eigenvalues are significant and how do we interpret their corresponding eigenvectors?

A hand-waving practical approach

- Suppose we find the values of σ and Q that best fit the bulk of the eigenvalue spectrum. We find

$$\sigma = 0.73; \; Q = 2.90$$

and obtain the following plot:



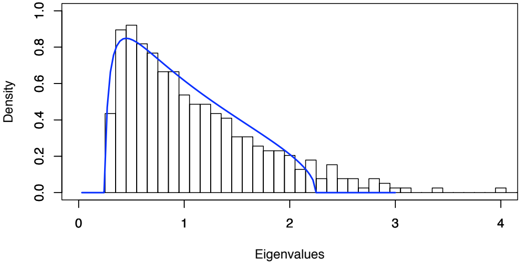
- Maximum and minimum Marčenko-Pastur eigenvalues are 1.34 and 0.09 respectively. Finiteness effects could take the maximum eigenvalue to 1.38 at the most.

Some analysis

- If we are to believe this estimate, a fraction $\sigma^2 = 0.53$ of the variance is explained by eigenvalues that correspond to random noise. The remaining fraction 0.47 has information.
- From the plot, it looks as if we should cut off eigenvalues above 1.5 or so.
- Summing the eigenvalues themselves, we find that 0.49 of the variance is explained by eigenvalues greater than 1.5
- Similarly, we find that 0.47 of the variance is explained by eigenvalues greater than 1.78
- The two estimates are pretty consistent!

More carefully: correlation matrix of residual returns

- Now, for each stock, subtract factor returns associated with the top 25 eigenvalues ($\lambda > 1.6$).
- We find that $\sigma = 1$; $Q = 4$ gives the best fit of the Marčenko-Pastur density and obtain the following plot:



- Maximum and minimum Marčenko-Pastur eigenvalues are 2.25 and 0.25 respectively.

Distribution of eigenvector components

- If there is no information in an eigenvector, we expect the distribution of the components to be a maximum entropy distribution.
- Specifically, if we normalized the eigenvector **u** such that its components u_i satisfy

$$\sum_i^M u_i^2 = M,$$

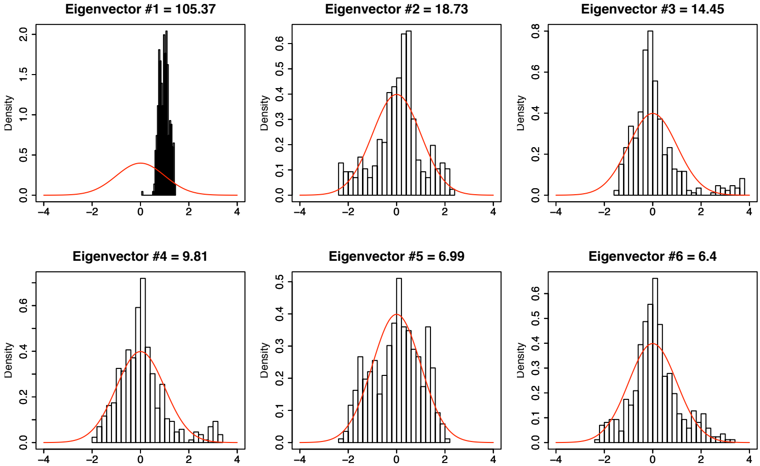
the distribution of the u_i should have the limiting density

$$p(u) = \sqrt{\frac{1}{2\,\pi}} \exp\left\{-\frac{u^2}{2}\right\}$$

- Let's now superimpose the empirical distribution of eigenvector components and the zero-information limiting density for various eigenvalues.

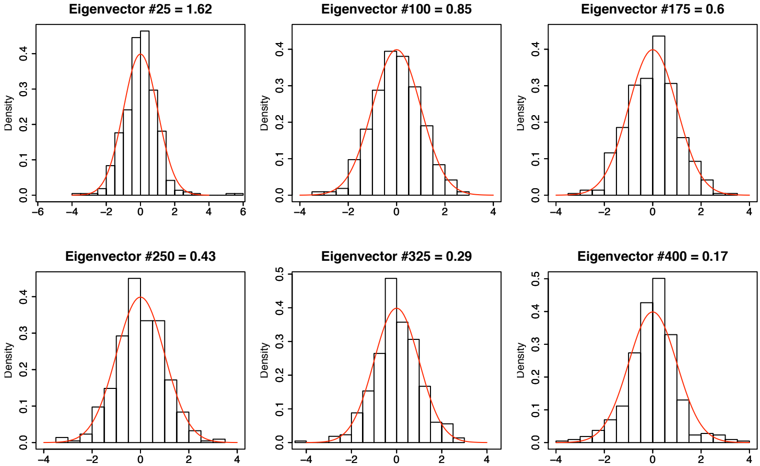
Informative eigenvalues

Here are pictures for the six largest eigenvalues:



Non-informative eigenvalues

Here are pictures for six eigenvalues in the bulk of the distribution:



The resulting recipe

1. Fit the Marčenko-Pastur distribution to the empirical density to determine Q and σ.

The resulting recipe

1. Fit the Marčenko-Pastur distribution to the empirical density to determine Q and σ.
2. All eigenvalues above some number λ^* are considered informative; otherwise eigenvalues relate to noise.

The resulting recipe

1. Fit the Marčenko-Pastur distribution to the empirical density to determine Q and σ.
2. All eigenvalues above some number λ^* are considered informative; otherwise eigenvalues relate to noise.
3. Replace all noise-related eigenvalues λ_i below λ^* with a constant and renormalize so that $\sum_{i=1}^{M} \lambda_i = M$.
   - Recall that each eigenvalue relates to the variance of a portfolio of stocks. A very small eigenvalue means that there exists a portfolio of stocks with very small out-of-sample variance – something we probably don't believe.

The resulting recipe

1. Fit the Marčenko-Pastur distribution to the empirical density to determine Q and σ.
2. All eigenvalues above some number λ^* are considered informative; otherwise eigenvalues relate to noise.
3. Replace all noise-related eigenvalues λ_i below λ^* with a constant and renormalize so that $\sum_{i=1}^{M} \lambda_i = M$.
   - Recall that each eigenvalue relates to the variance of a portfolio of stocks. A very small eigenvalue means that there exists a portfolio of stocks with very small out-of-sample variance – something we probably don't believe.
4. Undo the diagonalization of the sample correlation matrix $\mathbf{C}$ to obtain the denoised estimate $\mathbf{C}'$.

An extra detail

- In general, we will have $C'_{ii} \neq 1$.
- We can set diagonal elements to one by reweighting eigenvector components.
  - Let D be the diagonal matrix with elements $D_i = 1/\sqrt{C'_{ii}}$. Then

    $$C'' = D\, C'\, D$$

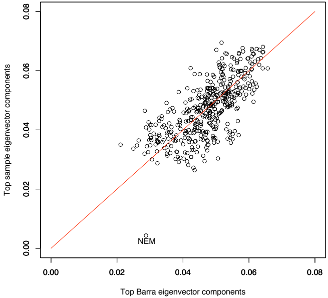
    has $C''_{ii} = 1$.
- Eigenvector components will be reweighted by $1/\sqrt{C_{ii}}$.

Comparison with Barra

- We might wonder how this random matrix recipe compares to Barra.
- For example:
  - How similar are the top eigenvectors of the sample and Barra matrices?
  - How similar are the eigenvalue densities of the filtered and Barra matrices?
  - How do the minimum variance portfolios compare in-sample and out-of-sample?

Comparing the top eigenvector

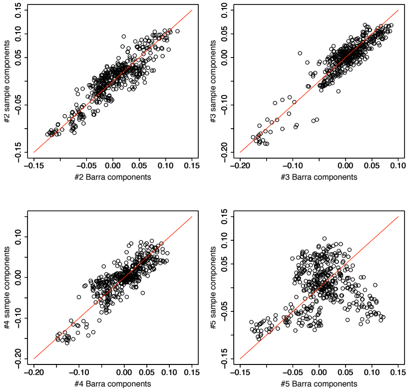
- We compare the eigenvectors corresponding to the top eigenvalue (the market components) of the sample and Barra correlation matrices:



- The eigenvectors are rather similar except for Newmont (NEM) which has no weight in the sample market component.

The next four eigenvectors

- The next four are:



- The first three of these are very similar but #5 diverges.

The minimum variance portfolio

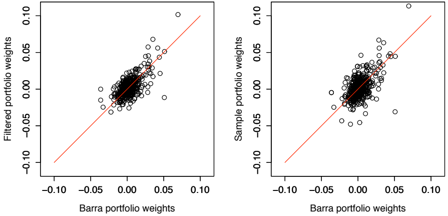
- We may construct a minimum variance portfolio by minimizing the variance $\mathbf{w}'.\mathbf{\Sigma}.\mathbf{w}$ subject to $\sum_i w_i = 1$.
- The weights in the minimum variance portfolio are given by

$$w_i = \frac{\sum_j \sigma_{ij}^{-1}}{\sum_{i,j} \sigma_{ij}^{-1}}$$

  where σ_{ij}^{-1} are the elements of $\mathbf{\Sigma^{-1}}$.
- We compute characteristics of the minimum variance portfolios corresponding to
    - the sample covariance matrix
    - the filtered covariance matrix (keeping only the top 25 factors)
    - the Barra covariance matrix

Comparison of portfolios

- We compute the minimum variance portfolios given the sample, filtered and Barra correlation matrices respectively.
- From the picture below, we see that the filtered portfolio is closer to the Barra portfolio than the sample portfolio.



- Consistent with the pictures, we find that the absolute position sizes (adding long and short sizes) are: Sample: 4.50; Filtered: 3.82; Barra: 3.40

In-sample performance

- In sample, these portfolios performed as follows:

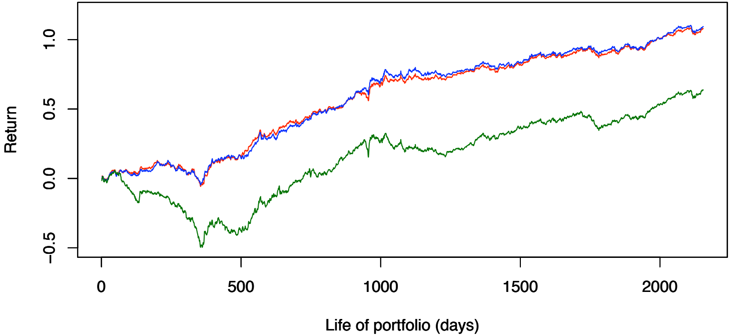


Figure: Sample in red, filtered in blue and Barra in green.

In-sample characteristics

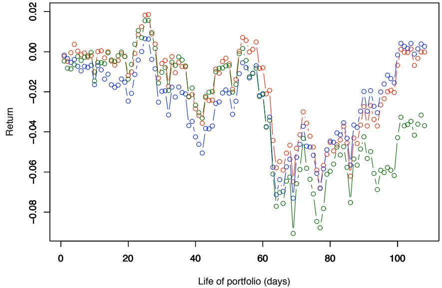
- In-sample statistics are:

	Volatility	Max Drawdown
Sample	0.523%	18.8%
Filtered	0.542%	17.7%
Barra	0.725%	55.5%

- Naturally, the sample portfolio has the lowest in-sample volatility.

Out of sample comparison

- We plot minimum variance portfolio returns from 04/26/2007 to 09/28/2007.
- The sample, filtered and Barra portfolio performances are in red, blue and green respectively.



- Sample and filtered portfolio performances are pretty similar and both much better than Barra!

Out of sample summary statistics

- Portfolio volatilities and maximum drawdowns are as follows:

	Volatility	Max Drawdown
Sample	0.811%	8.65%
Filtered	0.808%	7.96%
Barra	0.924%	10.63%

- The minimum variance portfolio computed from the filtered covariance matrix wins according to both measures!
  - However, the sample covariance matrix doesn't do too badly ...

Main result

- It seems that the RMT filtered sample correlation matrix performs better than Barra.
  - Although our results here indicate little improvement over the sample covariance matrix from filtering, that is probably because we had $Q = 5$.
  - In practice, we are likely to be dealing with more stocks (M greater) and fewer observations (T smaller).
- Moreover, the filtering technique is easy to implement.

When and when not to use a factor model

Quoting from Fan, Fan and Lv:

- The advantage of the factor model lies in the estimation of the inverse of the covariance matrix, not the estimation of the covariance matrix itself. When the parameters involve the inverse of the covariance matrix, the factor model shows substantial gains, whereas when the parameters involved the covariance matrix directly, the factor model does not have much advantage.

Moral of the story

Fan, Fan and Lv's conclusion can be extended to all techniques for "improving" the covariance matrix:

- In applications such as portfolio optimization where the inverse of the covariance matrix is required, it is important to use a better estimate of the covariance matrix than the sample covariance matrix.
  - Noise in the sample covariance estimate leads to spurious sub-portfolios with very low or zero predicted variance.
- In applications such as risk management where only a good estimate of risk is required, the sample covariance matrix (which is unbiased) should be used.

Miscellaneous thoughts/ observations

- There are reasons to think that the RMT recipe might be robust to changes in details:
  - It doesn't really seem to matter much exactly how many factors you keep.
  - In particular, Tracy-Widom seems to be irrelevant in practice.

Miscellaneous thoughts/ observations

- There are reasons to think that the RMT recipe might be robust to changes in details:
  - It doesn't really seem to matter much exactly how many factors you keep.
  - In particular, Tracy-Widom seems to be irrelevant in practice.
- The better performance of the RMT correlation matrix relative to Barra probably relates to the RMT filtered matrix uncovering real correlation structure in the time series data which Barra does not capture.

Miscellaneous thoughts/ observations

- There are reasons to think that the RMT recipe might be robust to changes in details:
  - It doesn't really seem to matter much exactly how many factors you keep.
  - In particular, Tracy-Widom seems to be irrelevant in practice.
- The better performance of the RMT correlation matrix relative to Barra probably relates to the RMT filtered matrix uncovering real correlation structure in the time series data which Barra does not capture.
- With $Q = 5$, the sample covariance does very well, even when it is inverted. That suggests that the key is to reduce sampling error in correlation estimates.
  - We could perhaps increase sample size by subsampling (hourly for example).

Miscellaneous thoughts/ observations

- There are reasons to think that the RMT recipe might be robust to changes in details:
  - It doesn't really seem to matter much exactly how many factors you keep.
  - In particular, Tracy-Widom seems to be irrelevant in practice.
- The better performance of the RMT correlation matrix relative to Barra probably relates to the RMT filtered matrix uncovering real correlation structure in the time series data which Barra does not capture.
- With $Q = 5$, the sample covariance does very well, even when it is inverted. That suggests that the key is to reduce sampling error in correlation estimates.
  - We could perhaps increase sample size by subsampling (hourly for example).
- It seems that correlations are really different at different timescales so we can't use tick data to get a better estimate of the one-day correlation matrix.

References

Jean-Philippe Bouchaud, Marc Mézard, and Marc Potters.
Theory of Financial Risk and Derivative Pricing: From Statistical Physics to Risk Management.
Cambridge University Press, Cambridge, 2 edition, 2003.

Jianqing Fan, Yingying Fan, and Jinchi Lv III.
Large dimensional covariance matrix estimation via a factor model.
Journal of Econometrics **147**(1) 186–197 (2008).

J.-P. Bouchaud G. Biroli and M. Potters.
On the top eigenvalue of heavy-tailed random matrices.
Europhysics Letters, 78(1):10001 (5pp), April 2007.

Iain M. Johnstone.
High dimensional statistical inference and random matrices.
In *International Congress of Mathematicians I* 307–333, Eur. Math. Soc., Zürich (2007).

Vasiliki Plerou, Parameswaran Gopikrishnan, Bernd Rosenow, Luís A. Nunes Amaral, Thomas Guhr, and H. Eugene Stanley.
Random matrix approach to cross correlations in financial data.
Physical Review E, 65(6):066126, June 2002.

Jean-Philippe Bouchaud and Marc Potters.
Financial Applications of Random Matrix Theory: a short review.
In *The Oxford Handbook of Random Matrix Theory*.
Oxford University Press, 2011.

Eugene P. Wigner.
Characteristic vectors of bordered matrices with infinite dimensions.
The Annals of Mathematics, 2nd Ser., 62(3):548–564, November 1955.

The Eigenvalue Trace Formula

For any real symmetric matrix $\mathbf{A}$, $\exists$ a unitary matrix $\mathbf{U}$ consisting of the (normalized) eigenvectors of A such that

$$\mathbf{L} = \mathbf{U}' \, \mathbf{A} \, \mathbf{U}$$

is diagonal. The entries λ_i of $\mathbf{L}$ are the eigenvalues of $\mathbf{A}$.
Noting that $\mathbf{L}^k = \mathbf{U}' \, \mathbf{A}^k \, \mathbf{U}$ it follows that the eigenvalues of $\mathbf{A}^k$ are λ_i^k. In particular,

$$\mathrm{Tr}\left[\mathbf{A}^k\right] = \mathrm{Tr}\left[L^k\right] = \sum_i^N \lambda_i^k \to N \, \mathbb{E}[\lambda^k] \text{ as } N \to \infty$$

That is, the kth moment of the distribution $\rho(\lambda)$ of eigenvalues is given by

$$\mathbb{E}[\lambda^k] = \lim_{N \to \infty} \frac{1}{N} \mathrm{Tr}\left[\mathbf{A}^k\right]$$

Matching moments

Then, to prove Wigner's semi-circle law, we need to show that the moments of the semicircle distribution are equal to the the traces on the right hand side in the limit $N \to \infty$.
For example, if $\mathbf{A}$ is a Wigner matrix,

$$\frac{1}{N} \operatorname{Tr}[\mathbf{A}] = \frac{1}{N} \sum_{i}^{N} a_{ii} \to 0 \text{ as } N \to \infty$$

and 0 is the first moment of the semi-circle density.
Now for the second moment:

$$\frac{1}{N} \operatorname{Tr}\left[\mathbf{A}^2\right] = \frac{1}{N} \sum_{i,j}^{N} a_{ij}\, a_{ji} = \frac{1}{N} \sum_{i,j}^{N} a_{ij}^2 \to \sigma^2 \text{ as } N \to \infty$$

It is easy to check that σ^2 is the second moment of the semi-circle density.

The third moment

$$\frac{1}{N}\operatorname{Tr}\left[\mathbf{A}^3\right] = \frac{1}{N}\sum_{i,j,k}^{N} a_{ij}\, a_{jk}\, a_{ki}$$

Because the a_{ij} are assumed *iid*, this sum tends to zero. This is true for all odd powers of $\mathbf{A}$ and because the semi-circle law is symmetric, all odd moments are zero.

The fourth moment

$$\frac{1}{N}\,\mathrm{Tr}\left[\mathbf{A}^4\right] = \frac{1}{N}\sum_{i,j,k,l}^{N} a_{ij}\,a_{jk}\,a_{kl}\,a_{li}$$

To get a nonzero contribution to this sum in the limit $N \to \infty$, we must have at least two pairs of indices equal. We also get a nonzero contribution from the N cases where all four indices are equal but that contribution goes away in the limit $N \to \infty$. Terms involving diagonal entries a_{ii} also vanish in the limit. In the case $k = 4$, we are left with two distinct terms to give

$$\frac{1}{N}\,\mathrm{Tr}\left[\mathbf{A}^4\right] = \frac{1}{N}\sum_{i,j,k,l}^{N} \left\{a_{ij}\,a_{ji}\,a_{il}\,a_{li} + a_{ij}\,a_{jk}\,a_{kj}\,a_{ji}\right\} \to 2\,\sigma^4 \text{ as } N \to \infty$$

Naturally, $2\,\sigma^4$ is the fourth moment $\mathbb{E}^{\rho}[\lambda^4]$ of the semi-circle density.

Higher moments

- Only products of pairs of matrix elements contribute.
  - Just as in the case of the fourth moment, there are at least a factor N fewer terms involving higher order combinations.
- So we need to count how many different combinations of two elements contribute in the $N \to \infty$ limit.
  - This turns out to be the number of planar (non-crossing) ways of connecting pairs of elements.
- For $k = 2j$, the number of such combinations is given by the Catalan numbers

$$C_j = \frac{(2j)!}{(j+1)!\,j!}.$$

- Thus, for $k = 2j$,

$$\lim_{N\to\infty} \frac{1}{N} \operatorname{Tr}\left[\mathbf{A}^k\right] = C_j\,\sigma^{2j},$$

identical to the moments of the Wigner semicircle density.

Remarks on Wigner's result

- The elements a_{ij} of $\mathbf{A}$ don't need to be normally distributed. In Wigner's original proof, $a_{ij} = \pm\nu$ for some fixed ν.
  - However, we do need the higher moments of the distribution of the a_{ij} to vanish sufficiently rapidly.
  - In practice, this means that if returns are fat-tailed, we need to be careful.
- The Wigner semi-circle law is like a Central Limit theorem for random matrices.