The quantspec package in R

Mark Beers

Introduction

- The quantspec package in R enables us to perform quantile based spectral analysis of time series quickly and easily. The general structure of the following slides is as follows.
  - Introduce and compare quantile based spectral analysis and traditional spectral analysis mathematically
  - S&P 500 returns example
  - EEG example
  - References

Represent Time Series as a Sum of Periodic Functions

$$x_t = \sum_{k=1}^{q} A_k cos(2\pi\omega_k t + \phi) = \sum_{k=1}^{q} A_k cos(\phi) cos(2\pi\omega_k t) - A_k sin(\phi) sin(2\pi\omega_k t)$$

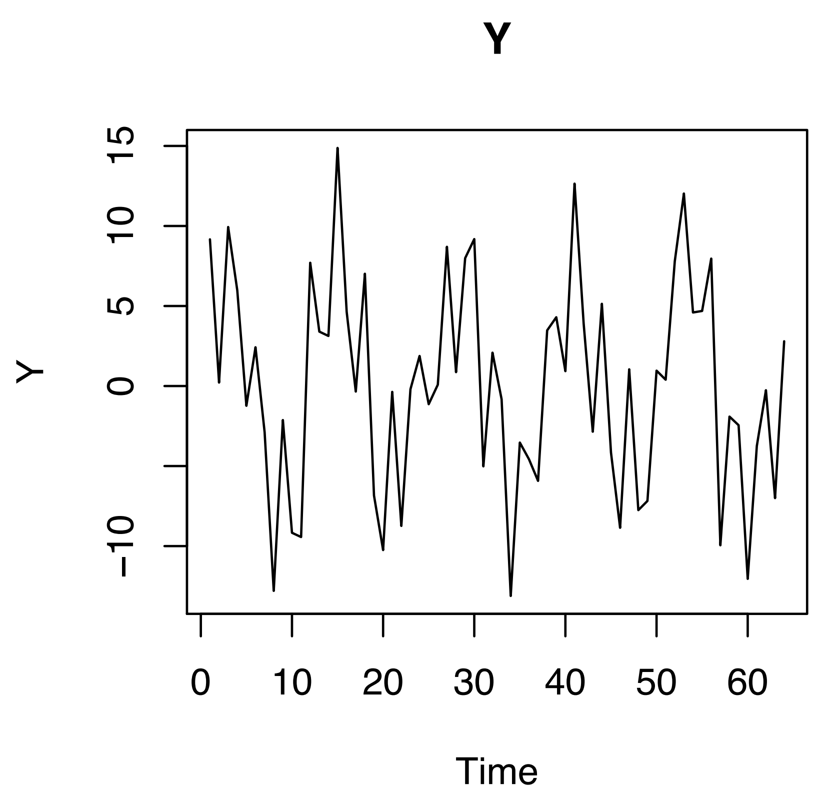
$$= \sum_{k=1}^{q} U_{1k} cos(2\pi\omega_k t) + U_{2k} sin(2\pi\omega_k t)$$

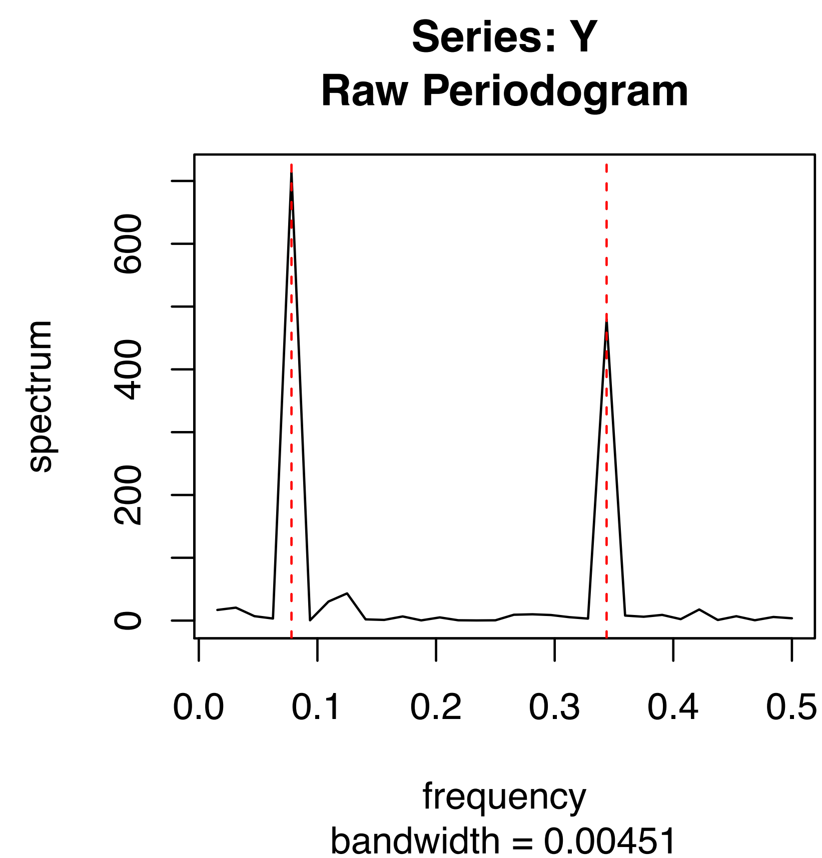
Traditional Spectral Analysis

Conceptual Introduction

Sometimes it's helpful to represent a time series in the frequency domain, meaning rather than as a series of successive measurements, we write the time series as a sum of cosines and sines at different frequencies. The most common graphical tool to represent which frequencies are present in the data is the periodogram.

$$Y = 3cos(2\pi\frac{5}{64}t) + 6sin(2\pi\frac{5}{64}t) + 2cos(2\pi\frac{22}{64}t) + 5sin(2\pi\frac{22}{64}t) + \epsilon, \quad \epsilon \sim N(0,4)$$





Periodogram Derivation

Harmonic Regression Approach

Traditional Spectral Analysis

$$(\beta_1, \beta_2) = \underset{\beta_1(j/n), \beta_2(j/n)}{argmin} \sum_{t=1}^{n} \left(x_t - \beta_1(0/n) - \sum_{j=1}^{\frac{n}{2}-1} \left[\beta_1(j/n) cos(2\pi \frac{j}{n} t) + \beta_2(j/n) sin(2\pi \frac{j}{n} t) \right] - \beta_1(1/2) cos(\pi t) \right)^2$$

$$Periodogram = I(j/n) = \frac{n}{4}(\beta_1(j/n)^2 + \beta_2(j/n)^2)$$

Quantile Spectral Analysis

$$\rho_\tau(u) = u(\tau - \mathbb{I}(u \leq 0))$$

$$Q_Y(\tau) = \underset{u}{argmin}\; E[\rho_\tau(Y - u)]$$

$$(\alpha, \beta_1(j/n), \beta_2(j/n)) = (\alpha, \beta_\tau) = \underset{\alpha, \beta_1, \beta_2}{argmin} \sum_{t=1}^{n} \rho_\tau \big(x_t - \alpha - \beta_1 cos(2\pi jt/n) - \beta_2 sin(2\pi jt/n) \big)$$

$$l(j/n, \tau_1, \tau_2) = \frac{n}{4} \beta_{\tau_1}(j/n)^T \begin{bmatrix} 1 & i \\ -i & 1 \end{bmatrix} \beta_{\tau_2}(j/n)$$

Periodogram Derivation

DFT Approach

Traditional Spectral Analysis

$$
\begin{aligned}
d(j/n) &= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} x_t e^{-2\pi i j t/n} \\
I(j/n) &= |d(j/n)|^2 \\
&= \sum_{h=-(n-1)}^{n-1} \hat{\gamma}(h) e^{-2\pi i j h/n}
\end{aligned}
$$

Quantile Spectral Analysis

$$
\begin{aligned}
r^q_{xx}(h,\tau_1,\tau_2) &= Cov(\mathbb{I}(x_t < Q_x(\tau_1)), \mathbb{I}(x_{t+h} < Q_x(\tau_2)) \\
\hat{r}^q_{xx}(h,\tau_1,\tau_2) &= \frac{1}{n} \sum (\tau_1 - \mathbb{I}(x_t < \hat{Q}_x(\tau_1))(\tau_2 - \mathbb{I}(x_{t+h} < \hat{Q}_x(\tau_2)) \\
L_{xx}(j/n,\tau_1,\tau_2) &= \frac{1}{2\pi} \sum_{h=-(n-1)}^{n-1} \hat{r}^q_{xx}(h,\tau_1,\tau_2) e^{-2\pi i j h/n} \\
&= \frac{1}{2\pi} d_{\tau_1}(-j/n) d_{\tau_2}(j/n) \\
\text{where} \quad d_\tau(j/n) &= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \mathbb{I}(x_t \leq \hat{Q}(\tau)) e^{-2\pi i j t/n}
\end{aligned}
$$

Notable Properties of Periodograms

- $I(j/n) = I((n-j)/n)$, so typically only show from $j/n = 0$ to $1/2$
- Traditional periodogram values can be interpreted as proportional to the correlation between your data and an arbitrary sine wave at frequency j/n. The quantile periodograms don't have this interpretation, but a higher quantile periodogram value still implies greater presence of that frequency in your data.
- In the traditional spectral analysis case, both approaches to deriving the periodogram yield identical results for most frequencies, whereas in quantile periodogram case, the periodograms generated by each approach converge as the number of samples increases.
- Neither the traditional periodogram nor the quantile periodogram is a consistent estimator of the spectral density.

Smoothing the Periodogram

To achieve a consistent estimator of the spectral density, we “smooth” or take a weighted average of the periodogram values in both traditional and quantile based spectral analysis.

$$B(\omega) = \{\lambda_s = 2\pi s/n : \omega - \frac{\pi}{2M} \leq \lambda_s \leq \omega + \frac{\pi}{2M}\}$$

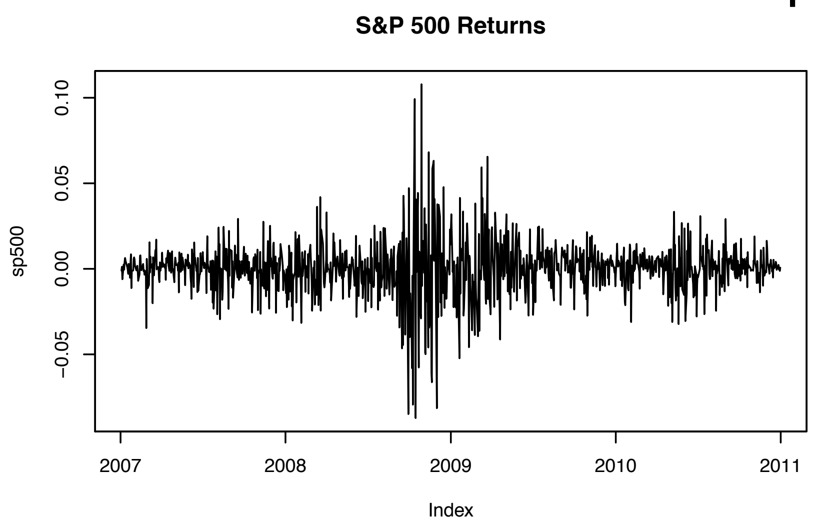
$$\hat{S}_{xx}(\omega, \tau_1, \tau 2) = \frac{1}{m} \sum_{\lambda_s \in B(\omega)} W(\lambda_s - \omega) \hat{L}_{xx}(\lambda_s, \tau_1, \tau 2)$$

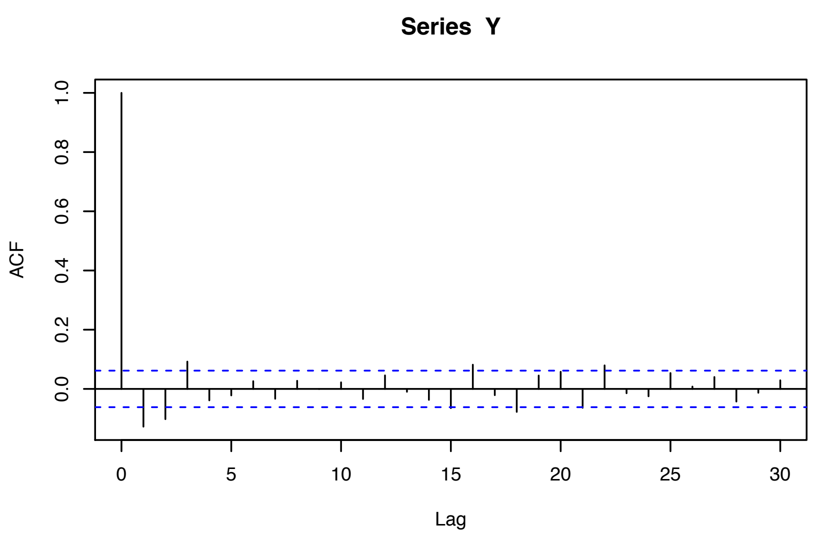
$$W(\lambda_s - \omega) = W(\omega - \lambda_s)$$

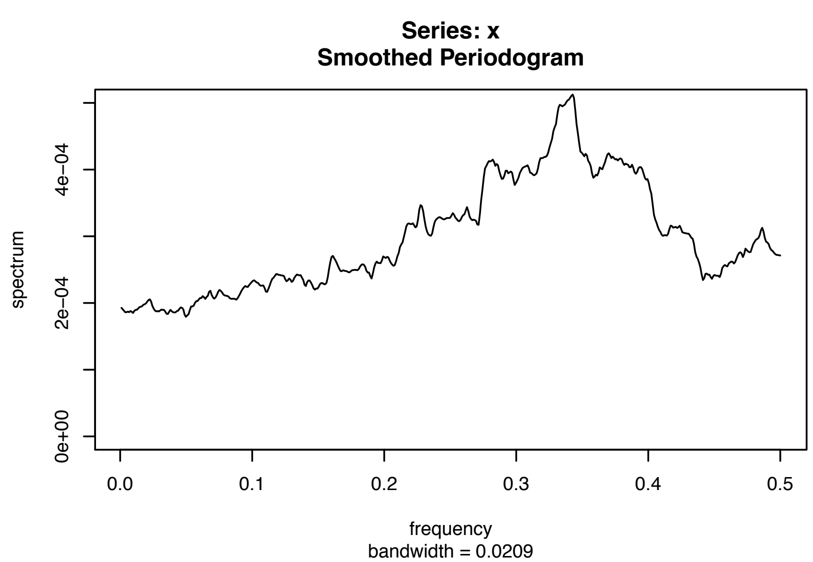
$$\sum_{\lambda_s \in B(\omega)} W(\lambda_s - \omega) = 1$$

- M is a parameter
- m is the number of Fourier Frequencies in B(w)
- The weights are symmetric about omega
- The sum of the weights must be one.
- As the smoothed periodogram is a consistent estimator, we should prefer to use it rather than the raw periodogram for making decisions about what frequencies are present in the data.

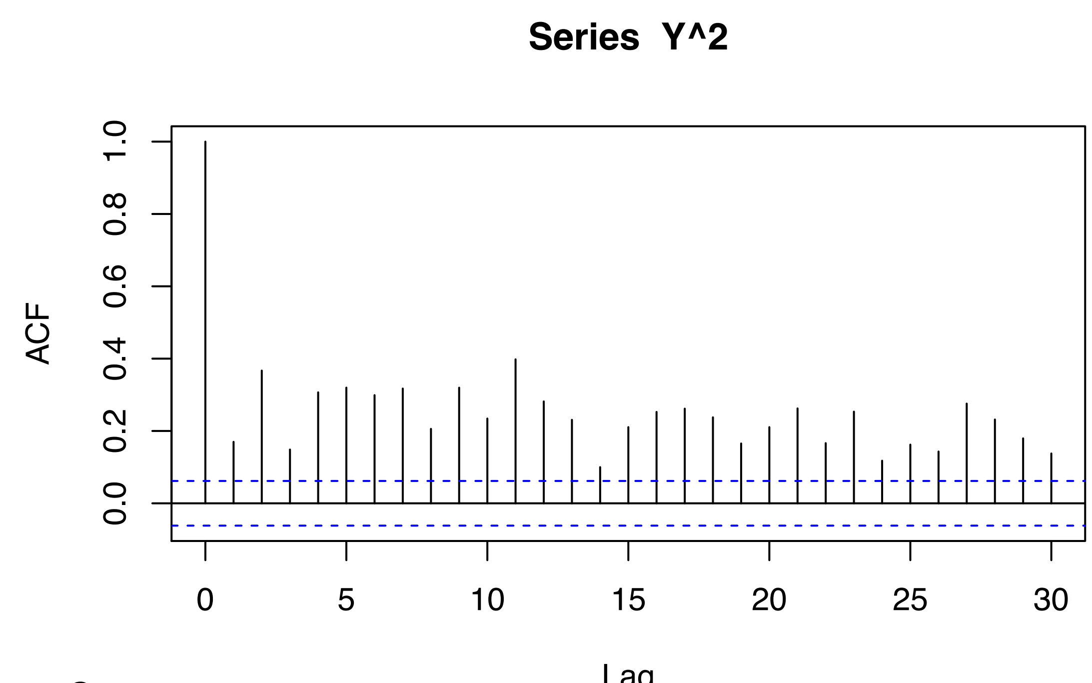
S&P 500 Returns Example
Exploratory Data Analysis



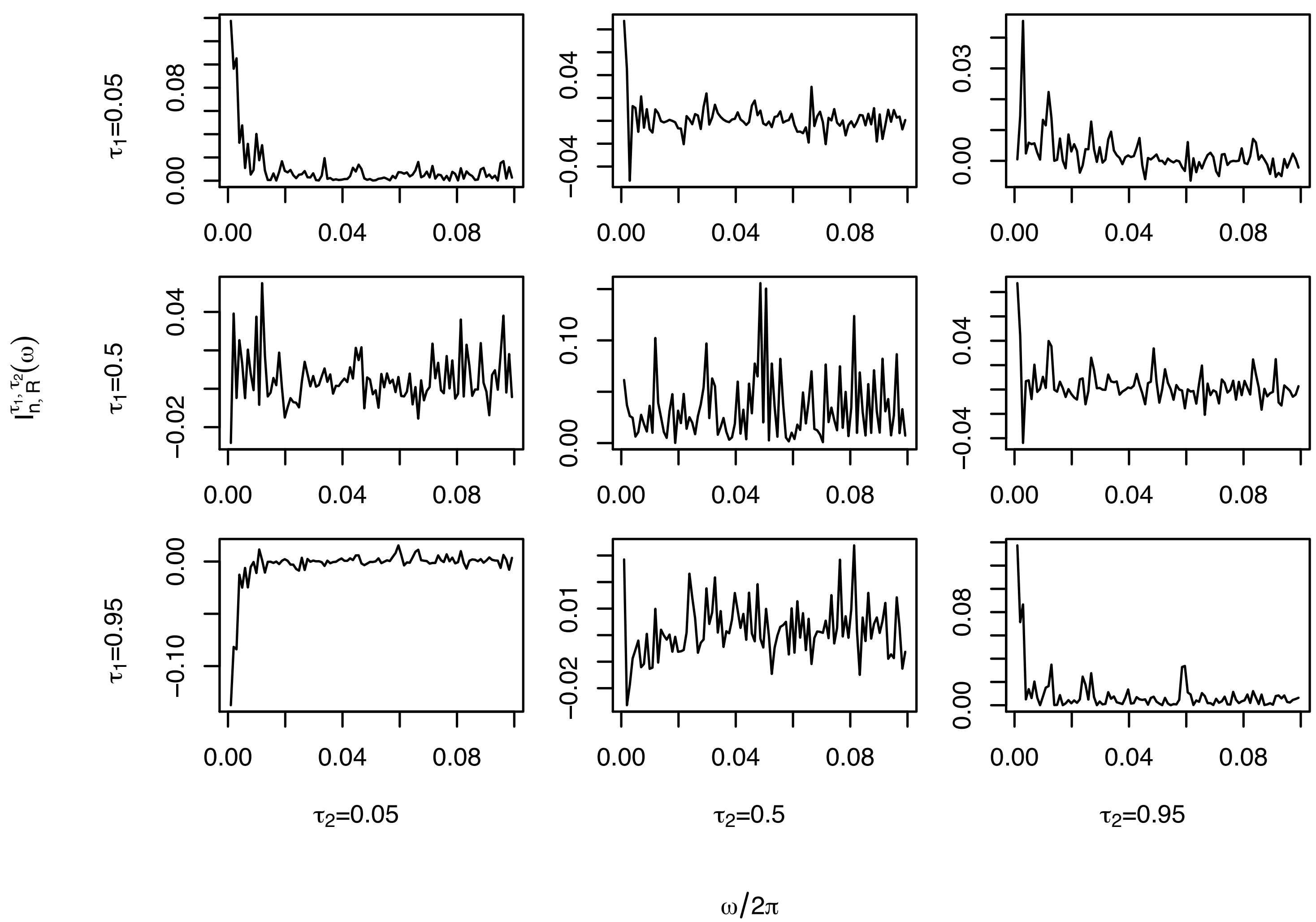




- Variance not constant over time, so non-stationary
- Nothing looks particularly significant in the autocorrelation plots, and since the periodogram is a transformation of the sample auto covariance, we expect nothing significant to pop up on the periodogram either, which is the case.
- However, we see significant autocorrelation in the squared series, which suggests that there is some dependence that isn't being captured by traditional spectral analysis.

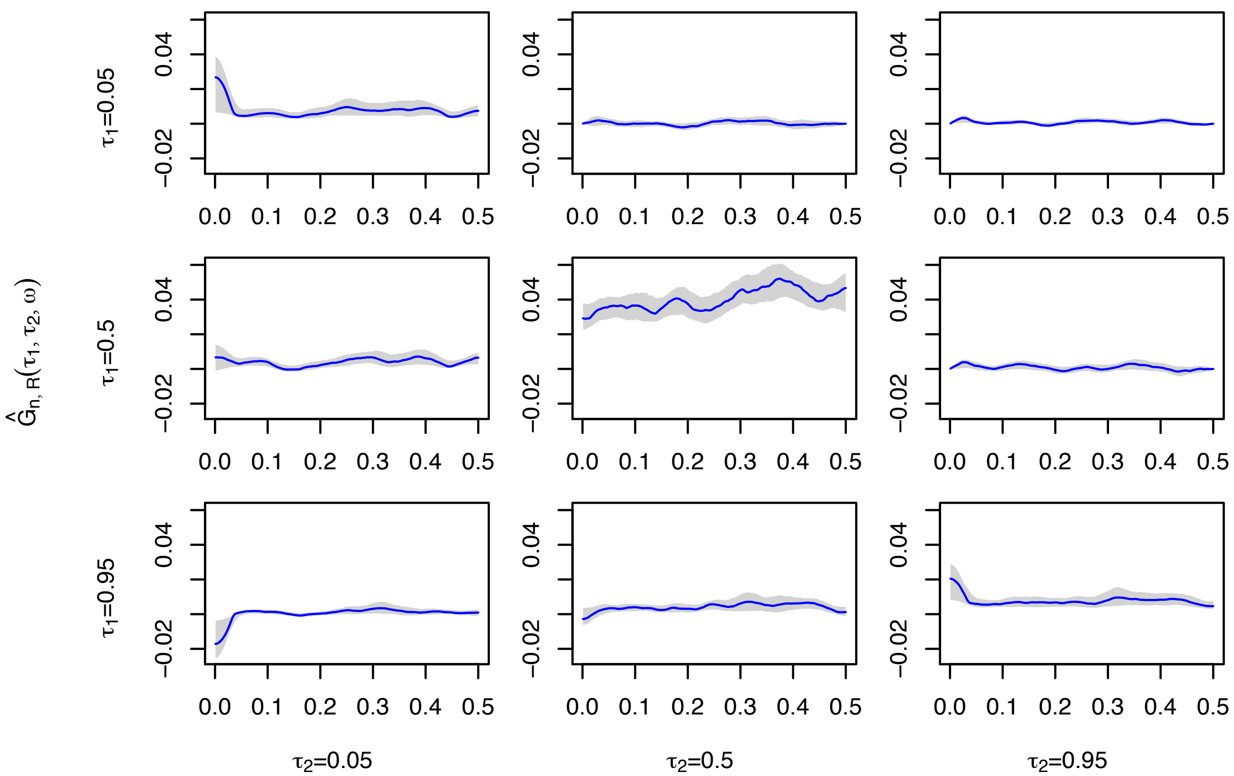


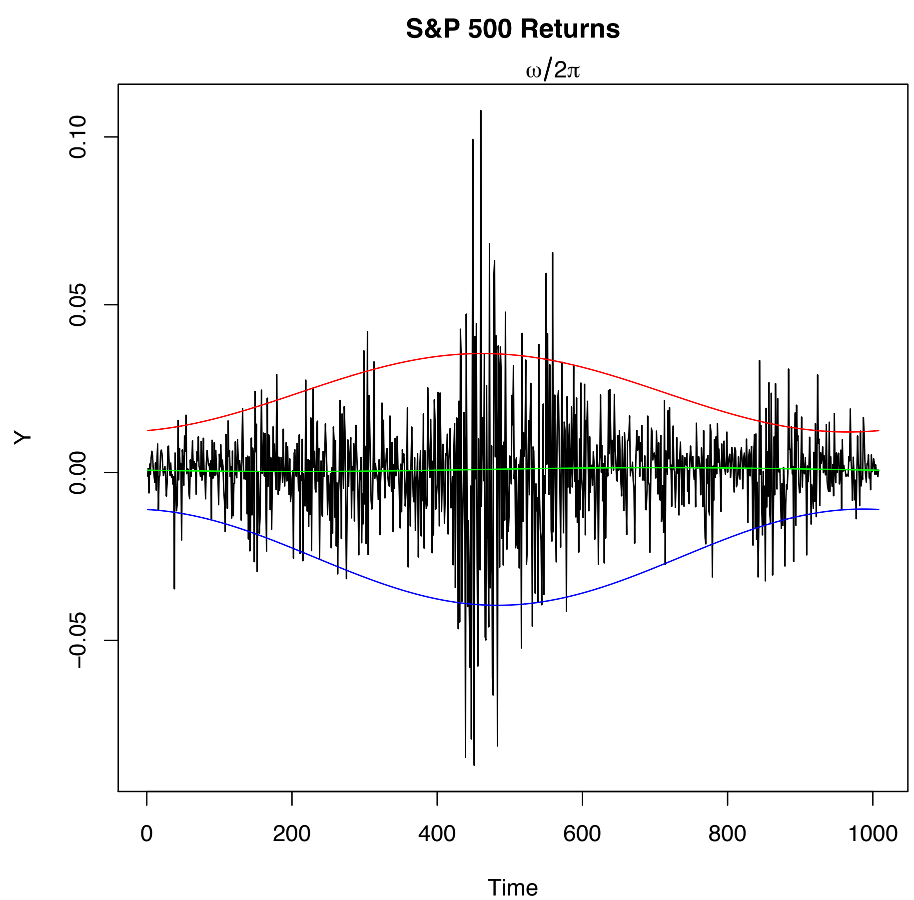
S&P 500 Returns Example
Raw Periodograms



- The main diagonal and below are real parts of the periodograms, while above the main diagonal are the imaginary parts of the periodogram.
- If tau1 = tau2, then the periodogram has no imaginary component.
- In the extreme quantiles 0.05 and 0.95, we note a significant spike at low frequencies
- These are the raw periodograms though, so look at smoothed periodograms for a better estimate of the spectral density.

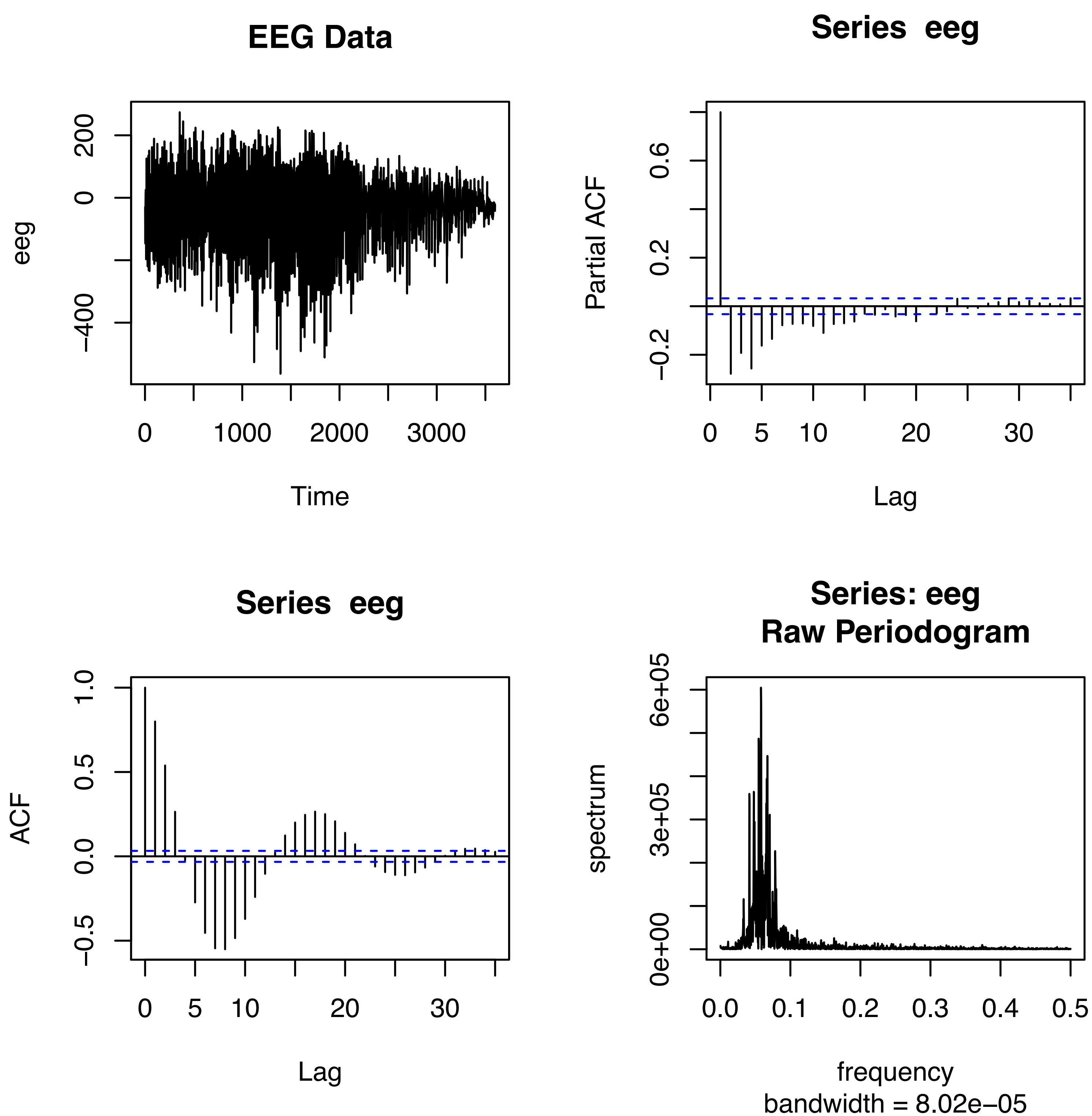
S&P 500 Returns Example
Smoothed Periodograms



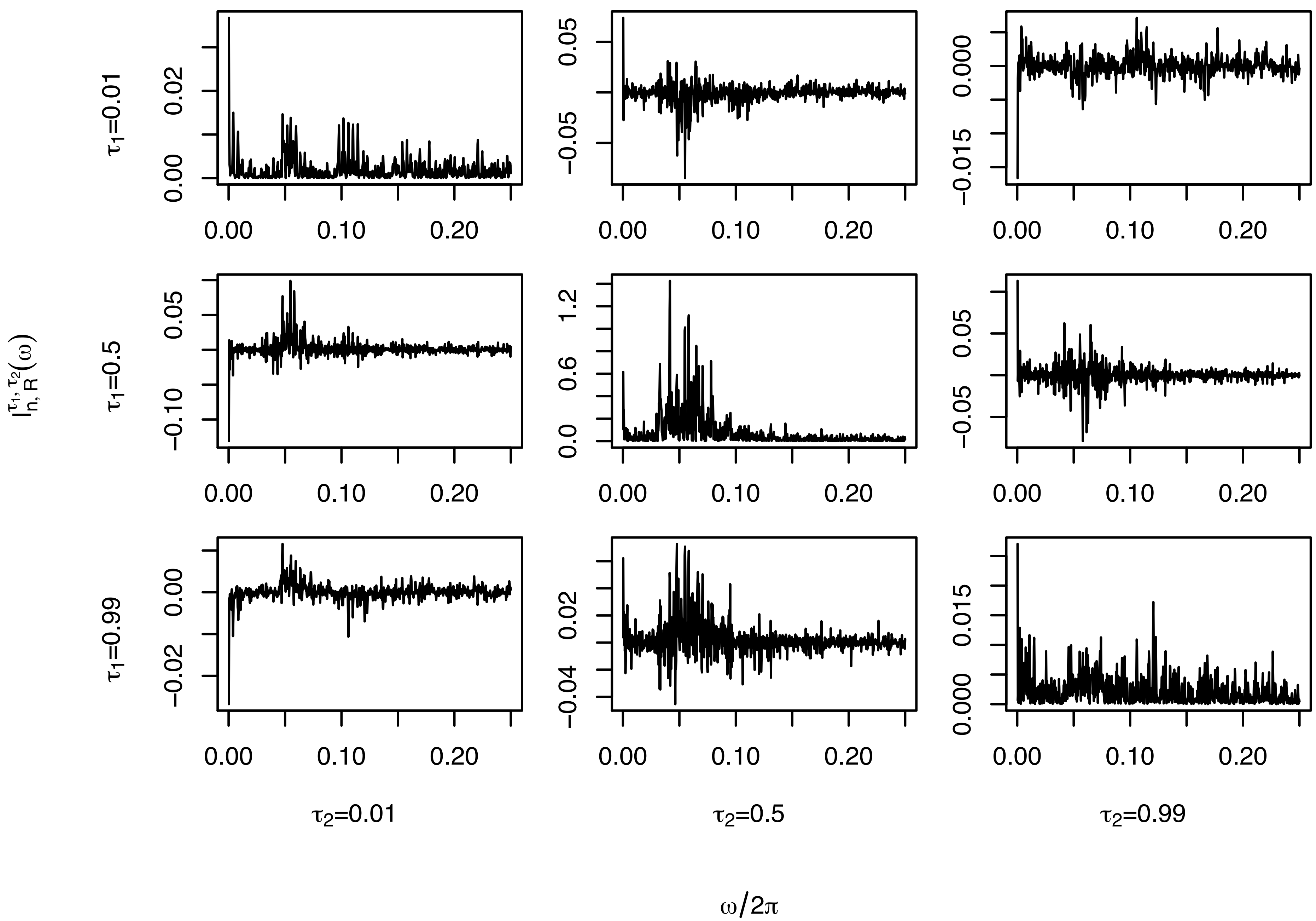


- We still see a tick up in lower frequencies in the extreme quantiles, but nothing significant in the median.
- The grey outline shows a confidence interval that the package can compute using a normal approximation or using a bootstrap approach
- Consider the frequency 1/1008, which is shown as being clearly greater than 0 in the (.05, 05) and (.95, .95) quantile periodograms. Getting the best fitting coefficients f=1/1008, we note that the reason the extreme quantile periodograms are high at 1/1008 is that the amplitude of the returns changes dramatically around t = 450.
- In contrast the traditional periodogram (green) finds almost no signal at frequency 1/1008, because the traditional periodogram uses least squares, which models the conditional mean.

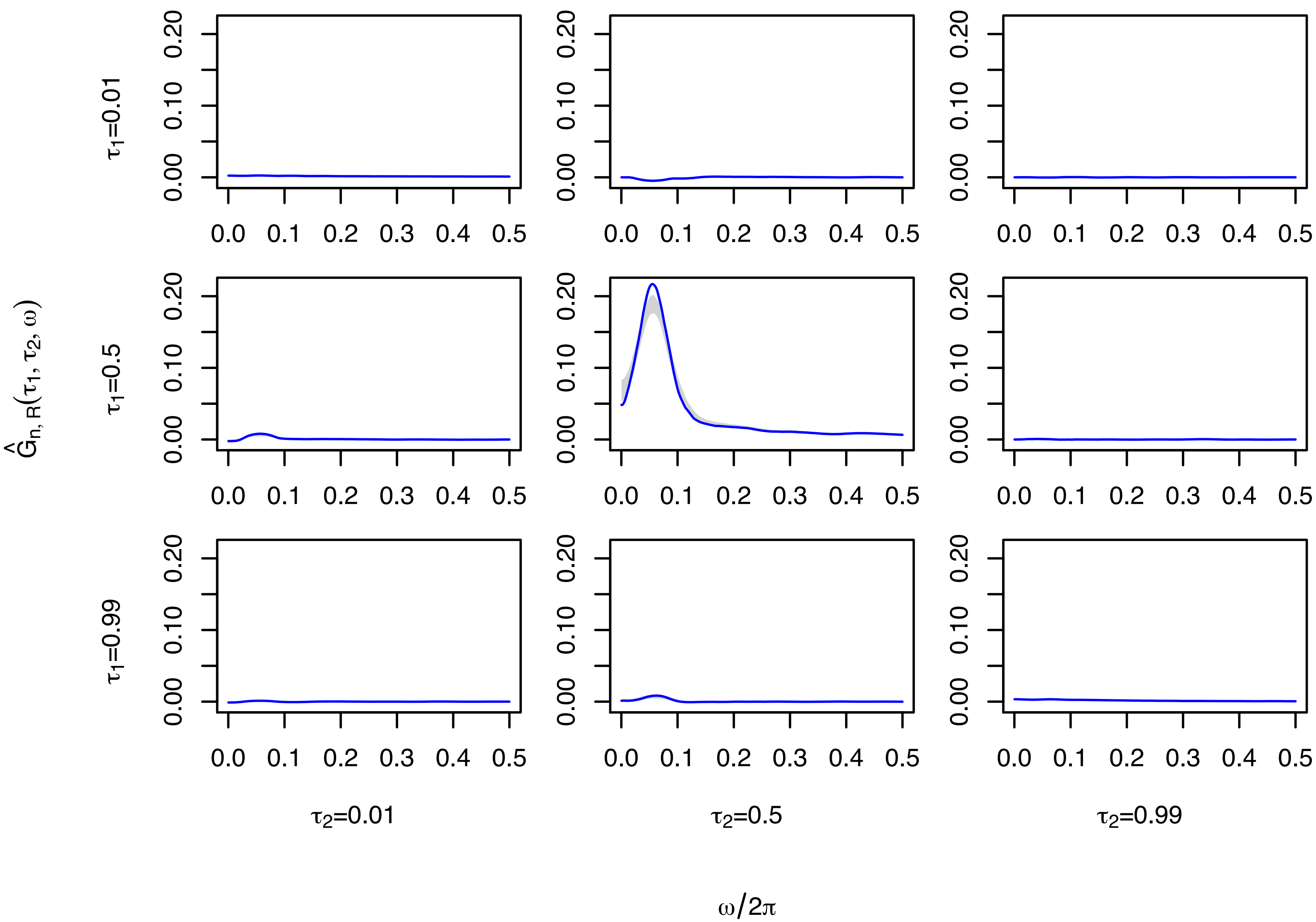
EEG Data EDA


EEG Data
eeg
200
0
−400
0
1000
2000
3000
Time
Series eeg
Partial ACF
0.6
0.2
−0.2
0
5
10
20
30
Lag
Series eeg
ACF
1.0
0.5
0.0
−0.5
0
5
10
20
30
Lag
Series: eeg
Raw Periodogram
spectrum
6e+05
3e+05
0e+00
0.0
0.1
0.2
0.3
0.4
0.5
frequency
bandwidth = 8.02e−05

EEG Raw Periodograms



EEG Smoothed Periodogram



Conclusion

- The quantile periodogram shares a very similar mathematical derivation to the traditional periodogram. In both cases, an intuitive harmonic regression derivation of the periodogram exists, while a more computationally efficient DFT derivation exists for real world use.

- The quantile periodogram has a similar interpretation to the traditional periodogram, with relatively large values at specific frequencies corresponding to strong presence of those frequencies in the data. This can be

- The quantile periodogram improves on the traditional periodogram though, in that it's more robust to outliers and as we saw in the S&P 500 case, enables us to pick up on periodic behavior in the extremes that may not show up in the center of the data.

References

- Robert H. Shumway and David S. Stoffer. (2005) Time Series Analysis and Its Applications. (Springer)

- Tobias Kley (2016) Quantile-Based Spectral Analysis in an Object-Oriented Framework and a Reference Implementation in R: The quantspec Package.

- Zhijie Zhao (Year) QAR and Quantile Time Series Analysis