Quantile Spectral Analysis

Mark Beers

October 17, 2019

1 Abstract

We introduce and subsequently compare and contrast traditional spectral analysis and quantile spectral analysis of time series data. Throughout the project we consider six datasets and use them to identify some of the shortcomings of traditional spectral analysis and to highlight the advantages of using quantile techniques. The six time series we consider are a simulated periodic dataset with non-stationarity in the mean, a time series of S&P 500 returns that exhibits non-stationarity in the variance, a recording from an EEG channel with non-stationary quasi-periodic behavior, two daily recordings of integrated vapor transport over three years, and finally a simulated white noise dataset to give a baseline idea of how various techniques respond to data with no autocorrelation. We estimate the spectral density of these six datasets using parametric, non-parametric and Bayesian approaches in both the traditional and quantile domains, observing consistency between the traditional approach and the quantile approach when we consider the median. While consistency is observed, the estimate of the spectral density that considers the median of the time series rather than the mean has some theoretical advantages, such as robustness to outliers. Further advantages of quantile approaches are demonstrated when we consider the extremes of our data. Obtaining an estimate of the spectral density of the 90th percentile of a time series, as is done in this project, is not possible using techniques from traditional spectral analysis and allows us to quantify important characteristics of a time series such as non-stationarity in the variance.

2 Introduction

Traditional spectral analysis is a well understood and useful tool, but has some limitations. In particular, the spectral density estimate provided by traditional spectral analysis is not robust to outliers, does not handle the heavier tails of non Gaussian processes well, and can not identify dependence in the extremes, among other weaknesses. To remedy this, modifications to traditional spectral analysis have been proposed. One such modification that addresses many of these shortcomings is quantile spectral analysis, which allows us to produce an estimate of the spectral density of a time series at an arbitrary quantile.

This project will introduce and discuss key concepts relevant to traditional spectral analysis, such as the autocorrelation function and the periodogram. Then we will consider a number of approaches to achieve an estimate of the spectral density, incuding smoothing the periodogram, computing the spectral density of a fitted AR model, and implementing a Bayesian approach using the BayesSpec package in R (Rosen et al., 2012). Next, very similar quantities in quantile spectral analysis will be introduced and computed, with an emphasis on highlighting the mathematical and practical

similarities between quantile and traditional spectral analysis. In particular, we will highlight the results of Xiao (2017) whose paper includes a good overview of quantile spectral analysis. It can be computationally burdensome to produce some of the quantities relevant to quantile spectral analysis, so this project will make heavy use of the quantspec package developed by Tobias Kley at Ruhr Universtat Bochum in Germany which provides methods for efficient computation (Kley, 2014).

List of Figures

3   Datasets

Throughout this project, we will reference a number of example data sets. We briefly provide a description of each dataset.

1. Periodic Data: A non-stationary dataset of $t = 1, 2, ..., 512$ observations from the following equation.

$$Y_t = 3\cos(2\pi\frac{5}{64}t) + 6\sin(2\pi\frac{5}{64}t) + 2\cos(2\pi\frac{22}{64}t) + 5\sin(2\pi\frac{22}{64}t) + \epsilon_t, \quad \epsilon_t \overset{i.i.d.}{\sim} N(0,4)$$

2. S&P 500 Returns: A dataset on S&P 500 returns from the quantspec package in R. The S&P 500 is a stock market index and this dataset details returns on investments in that index. The dataset consists of 1008 nearly daily observations of this quantity from January 2007 to January 2010.

3. EEG: A time series of 3600 observations corresponding to an electroencephalogram (EEG) recording on a patient who received electroconvulsive therapy. This EEG time series was recorded from a channel located at the center of the scalp (West et al., 1999).

4. IVT 1982-1985 & IVT 2012-2015: Two datasets measuring integrated vapor transport (IVT) over a region of California in the time periods 1982-1985 and 2012-2015. IVT is a quantity that measures changing levels of water vapor in the atmosphere, and is useful for quantifying drought or particularly rainy seasons. Both datasets include daily observations of this quantity. The two time periods chosen are of interest because 1982-1985 was a particularly rainy period, while 2012-2015 was a particularly dry period (Guan and Waliser, 2015).

5. White Noise: One thousand observations of white noise, simulated according to the following equation.

$$Y_t \overset{i.i.d.}{\sim} N(0,25)$$

The datasets are pictured in Figure 1. We will apply and compare different spectral methods to these data, including quantile spectral methods.
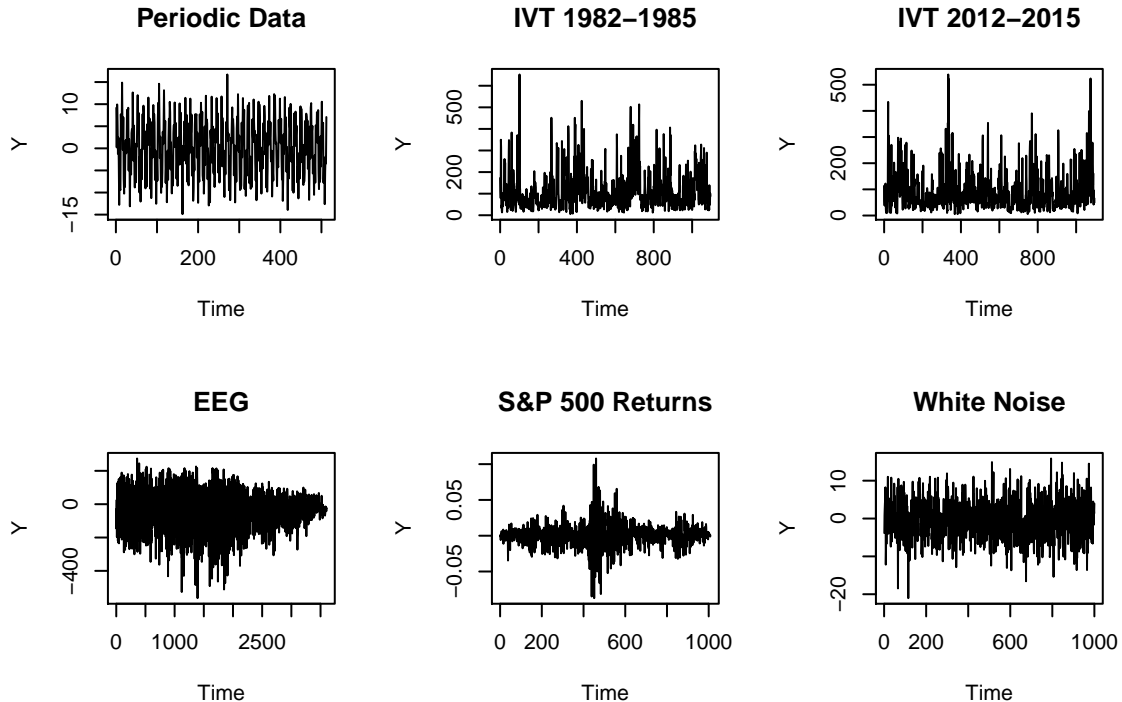
Figure 1: Datasets

4 Traditional Spectral Analysis

In this section, we will provide a review of traditional spectral analysis, which amounts to analysis of a time series in the frequency domain. The ability to transition from the time domain to the frequency domain when describing a time series has a number of advantages. In particular, switching to the frequency domain nearly eliminates dependence between adjacent datapoints. That is, the contribution of frequency $\frac{2\pi k}{n}$ in the data is roughly independent of the contribution of frequency $\frac{2\pi(k+1)}{n}$. This independence makes some computations much easier in the frequency domain. Another advantage of this transition is a potentially more compact and interpretable description of the time series. These points justify the importance of frequency domain approaches. However, there are some foundational concepts that are relevant to both approaches and important for this project. For instance, in the time domain, a common tool for trying to determine the presence of a linear relationship between the time series and lagged values of that time series is the autocorrelation function, or ACF. Given a stationary process $\{\mathbf{X}_t\}$, the autocovariance function γ and autocorrelation function ρ are defined below. Also defined below are sample estimates of the autocovariance function and autocorrelation function, labeled $\hat{\gamma}$ and $\hat{\rho}$, respectively. These estimates are based on n observed time series data points, x_t, for $t = 1, ..., n$. In both the theoretical and sample versions of these quantites, h represents an integer lag. See Shumway and Stoffer (2017) for examples and

4

properties of these functions.

$$\gamma(h) = cov(\mathbf{X}_{t+h}, \mathbf{X}_t), \qquad\qquad \hat{\gamma}(h) = \frac{1}{n}\sum_{t=1}^{n-h}(x_{t+h} - \bar{x})(x_t - \bar{x}),$$

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)}, \qquad\qquad\qquad \hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}.$$

In Figure 2, we have the sample ACF's plotted for each of the datasets shown in the previous section. We note that the ACF tails off for large lags in all cases but the purely periodic process. There is also some evident periodic behavior in the EEG data. Regarding the IVT datasets, we observe some gradual decay as the lag increases, but the ACF is always positive. Finally, we observe that the ACF of the S&P 500 returns data looks very similar to the ACF of the white noise data, with no autocorrelation being particularly strong. This is despite the fact that when we refer to Figure 1, we note that the white noise dataset is dramatically different from the S&P 500 Returns data. This suggests that the ACF does not fully capture the characteristics of time series processes, and in this project we will explore techniques that have the potential to give us richer insight.
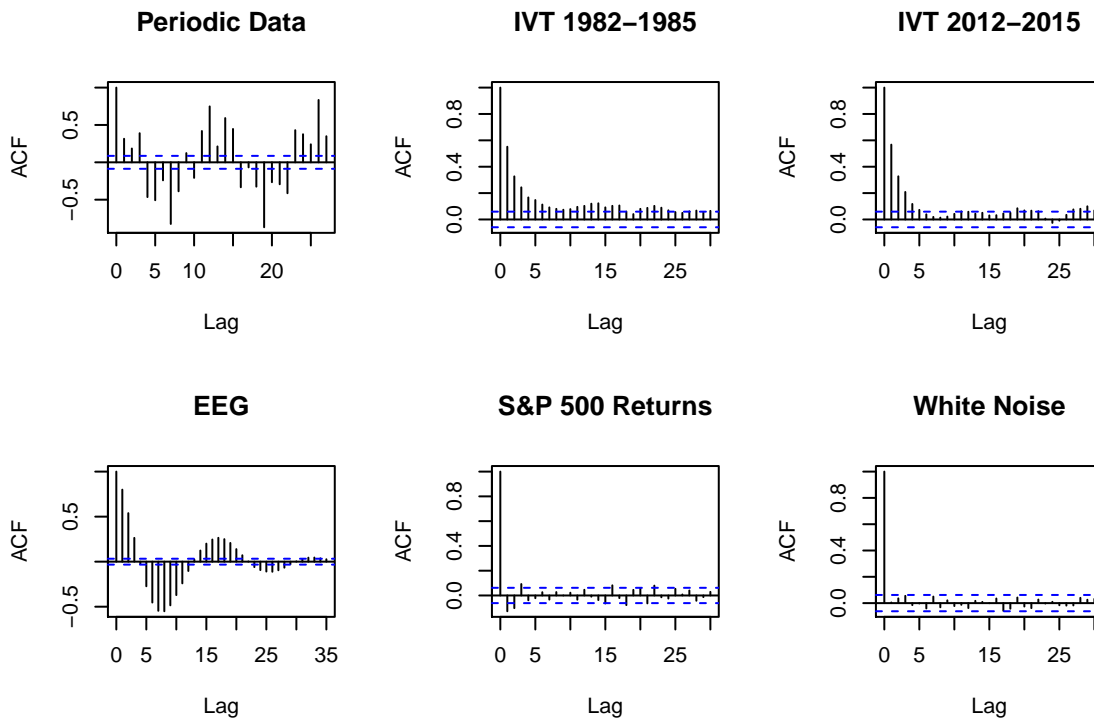

Figure 2: Sample Autocorrelation Functions

Spectral analysis consists of making inferences on a function called the spectral density. The spectral density $f(\omega)$ for a stationary process $\{\mathbf{X}_t\}$ is defined as the Fourier transform of the autocovariance function $\gamma(h)$ (Shumway and Stoffer, 2017):

$$f(\omega) = \sum_{h=-\infty}^{\infty} \gamma(h)e^{-i\omega h}, \quad -\pi \leq \omega \leq \pi.$$

5

The spectral density summarizes the frequency content of the data. If the spectral density is large for a particular frequency, then that suggests that such frequency contributes strongly to the behavior of our time series. One obvious way to estimate the spectral density would be to substitute in the sample autocovariance function for the true autocovariance function and then truncate the sum at a fixed number of lags. There are a couple of issues with this approach. First, this approach is clearly not going to characterize the processes in a very complete way, because we would be using a truncated version of the sample ACF that does not fully characterize the process. Second, it is not immediately apparent how the Fourier transform of the autocovariance function will yield a quantity that tells us how strongly present a frequency is in our data.

Next we present an alternative method of estimating the spectral density. We can express an arbitrary sine wave of frequency ν and phase ϕ in the following way, where $U_1 = A\cos(\phi)$ and $U_2 = A\sin(\phi)$ (Shumway and Stoffer, 2017):

$$A\sin(2\pi\nu t + \phi) = A\sin(2\pi\nu t)\cos(\phi) + A\cos(2\pi\nu t)\sin(\phi) = U_1\sin(2\pi\nu t) + U_2\cos(2\pi\nu t).$$

Therefore, for a specific frequency, we can estimate a sine wave of arbitrary phase and amplitude by estimating U_1 and U_2. Suppose we fixed the frequency ν and used a least squares approach to determine U_1 and U_2. If that fixed frequency is not a strong contributor to the behavior of the data, then we would expect the magnitudes of U_1 and U_2 to be small. However, if the frequency is a strong contributor, then we would expect significantly non-zero U_1 or U_2. So the size of U_1 and U_2 is related to the presence of a frequency in the data, and therefore it is important to get estimates for U_1 and U_2 at a wide range of frequencies. We get the estimates in the following way (Shumway and Stoffer, 2017):

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

In the above equation, $\beta_1(j/n)$ and $\beta_2(j/n)$ are coefficients of the cosine and sine covariates in the regression, each covariate at the Fourier frequency $2\pi j/n$, where j ranges from $1, ..., \frac{n}{2} - 1$. We note this is least squares regression, meaning that the resulting fitted values will provide an estimate for the mean of the time series. The contribution of a frequency to the mean of the time series can be found by looking at the periodogram. The value of the periodogram at the specific Fourier frequency $2\pi j/n$ is referred to as a periodogram ordinate and is denoted $I(j/n)$. $I(j/n)$ is defined as a function of the coefficients $\beta_1(j/n)$ and $\beta_2(j/n)$ and the specific form the coefficients $\beta_1(j/n)$ and $\beta_2(j/n)$ take after the minimization is given below (Shumway and Stoffer, 2017):

$$\beta_1(j/n) = \frac{2}{n}\sum_{t=1}^{n} x_t\cos(2\pi\omega_j t)$$

$$\beta_2(j/n) = \frac{2}{n}\sum_{t=1}^{n} x_t\sin(2\pi\omega_j t)$$

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

The periodogram has the property that each ordinate $I(j/n)$ is proportional to the correlation between an arbitrary sine wave at the frequency $2\pi j/n$ and the data.

Referring back to the minimization equation defined earlier, we note that we are regressing n data points onto n variables, so we are guaranteed a perfect fit. While in a normal regression setting we would want the number of observations to be much greater than the number of predictors, in this case we want to check for the presence of as many frequencies as possible while still getting a unique solution to the minimization equation. So, we perform this saturated regression with the hope that just about any periodic component present in the data will be identified. With a saturated regression, we have some computational issues when n is large, as we will have to invert a large $n \times n$ matrix. Fortunately however, there exists a computationally efficient method based on the Discrete Fourier Transform (DFT) for computing the periodogram that gives the exact same estimates as the regression approach defined above. Given our data, we can define the Discrete Fourier Transform $d(\omega_j)$ and the periodogram ordinates $I(\omega_j)$ as follows, where $i = \sqrt{-1}$ (Shumway and Stoffer, 2017):

$$d(\omega_j) = n^{-1/2} \sum_{t=1}^{n} x_t e^{-2\pi i \omega_j t}$$
$$= n^{-1/2} \sum_{t=1}^{n} x_t [\cos(2\pi \omega_j t) - i \sin(2\pi \omega_j t)]$$
$$I(\omega_j) = |d(\omega_j)|^2$$
$$j = 0, 1, ..., n-1, \quad \omega_j = j/n.$$

The following proof demonstrates that this DFT approach to computing the periodogram can be represented as something that looks very similar to the formula we use to compute the spectral density (Shumway and Stoffer, 2017):

$$d(\omega_j) = n^{-1/2} \sum_{t=1}^{n} x_t e^{-2\pi i \omega_j t} = n^{-1/2} \sum_{t=1}^{n} (x_t - \bar{x}) e^{-2\pi i \omega_j t}$$
$$I(\omega_j) = |d(\omega_j)|^2 = n^{-1} \sum_{t=1}^{n} \sum_{s=1}^{n} (x_t - \bar{x})(x_s - \bar{x}) e^{-2\pi i \omega_j (t-s)}$$
$$= n^{-1} \sum_{h=-(n-1)}^{n-1} \sum_{t=1}^{n-|h|} (x_{t+|h|} - \bar{x})(x_t - \bar{x}) e^{-2\pi i \omega_j h}$$
$$= \sum_{h=-(n-1)}^{n-1} \hat{\gamma}(h) e^{-2\pi i \omega_j h}.$$

The first line relies on the fact that $\sum_{t=1}^{n} e^{-2\pi i \omega_j t} = 0$. The third line is a reindexing that allows us to write the periodogram as a transformation of the sample autocovariance, meaning that the periodogram is just another representation of the information contained in the sample autocovariance.

In Figure 3 we provide periodograms for all datasets considered in this project plotted on the log scale. Note that these plots only range in frequency from 0 to 0.5, while our computation of the periodogram using the DFT approach considers frequencies up to $\frac{n-1}{n} > 0.5$ if $n > 2$. We do not plot frequencies greater than $1/2$ because the periodogram is symmetric about $1/2$, so looking at higher frequencies reveals no new information.
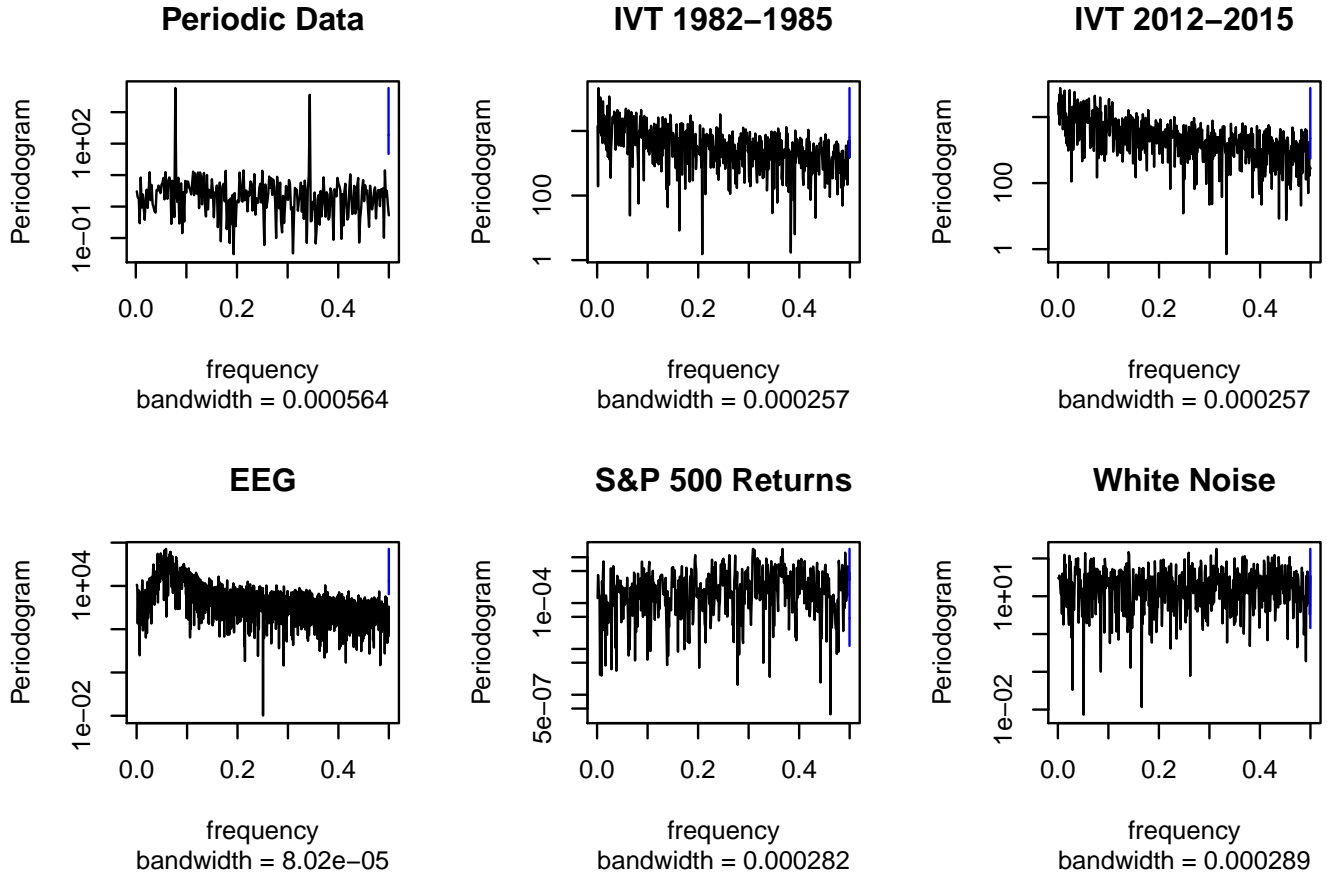
Figure 3: Sample Periodograms

In the periodic data, the spikes at 5/64 and 22/64 inform us of the presence of those frequencies in our data, so the periodogram accurately recovered the properties of the signal in this case. Looking at the EEG data, we note the periodogram identified the dominant frequency around $f = 0.07$, which corresponds roughly to 5Hz in the original data (the sampling rate for the EEG data considered here is 256/5 Hz). As discussed, the periodogram can be expressed as a transformation of the sample ACF. In the case of the S&P 500 returns time series, the ACF picked up nothing significant, and so we do not expect the periodogram to identify anything particularly interesting either. Indeed, we note that the periodogram for that dataset does not have any obvious spikes. Finally, we consider the IVT datasets. In both of these, low frequency components are detected, but high frequency components are not.

Though each of these periodograms is different, they all share one common feature, and that is what looks to be a lot of noise. This leads us to one of the main shortcomings of the periodogram as an estimate of the spectral density. Namely, the periodogram is not a consistent estimator of the spectral density meaning that even in the limit of infinite data, we will not approach the true spectral density. Instead, the periodogram ordinates approach constant non-zero variance, rather than zero variance as sample size increases. There are a number of ways to address this shortcoming. One way we can achieve a consistent estimator is to smooth or average the periodogram. Mathematically

this works as follows (Shumway and Stoffer, 2017):

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

$$\hat{S}_{xx}(\omega, \tau_1, \tau_2) = \frac{1}{m} \sum_{\lambda_s \in B(\omega)} W(\lambda_s - \omega) I(\lambda_s)$$

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

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

Above, $B(\omega)$ is a set of Fourier frequencies that we choose to average over, and M is a quantity that defines how large the averaging window is, and m is the cardinality of the set $B(\omega)$. Once we have selected the size of the averaging window, we perform a weighted average of the periodogram values in that window to compute $\hat{S}_{xx}$. There are a few conditions on the weights W. In particular, the weights must be symmetric about ω and they must sum up to 1. If these conditions are met, we will have a consistent estimator of the true spectral density. While smoothing the periodogram is a popular approach to achieve a consistent estimator of the true spectral density, other methods exist. For instance, one could also achieve the same results in the limit of infinite data by using a kernel estimator (Xiao, 2017), which performs a weighted average of the autocovariances in the DFT, rather than computing the periodogram ordinates and then averaging adjacent ordinates. While both approaches work, throughout this project any smoothed periodogram will be computed by averaging adjacent periodogram ordinates. In particular, Figure 4 shows smoothed periodograms using a Daniell kernel with width 9, which is an equally weighted average of 9 sequential periodogram ordinates.
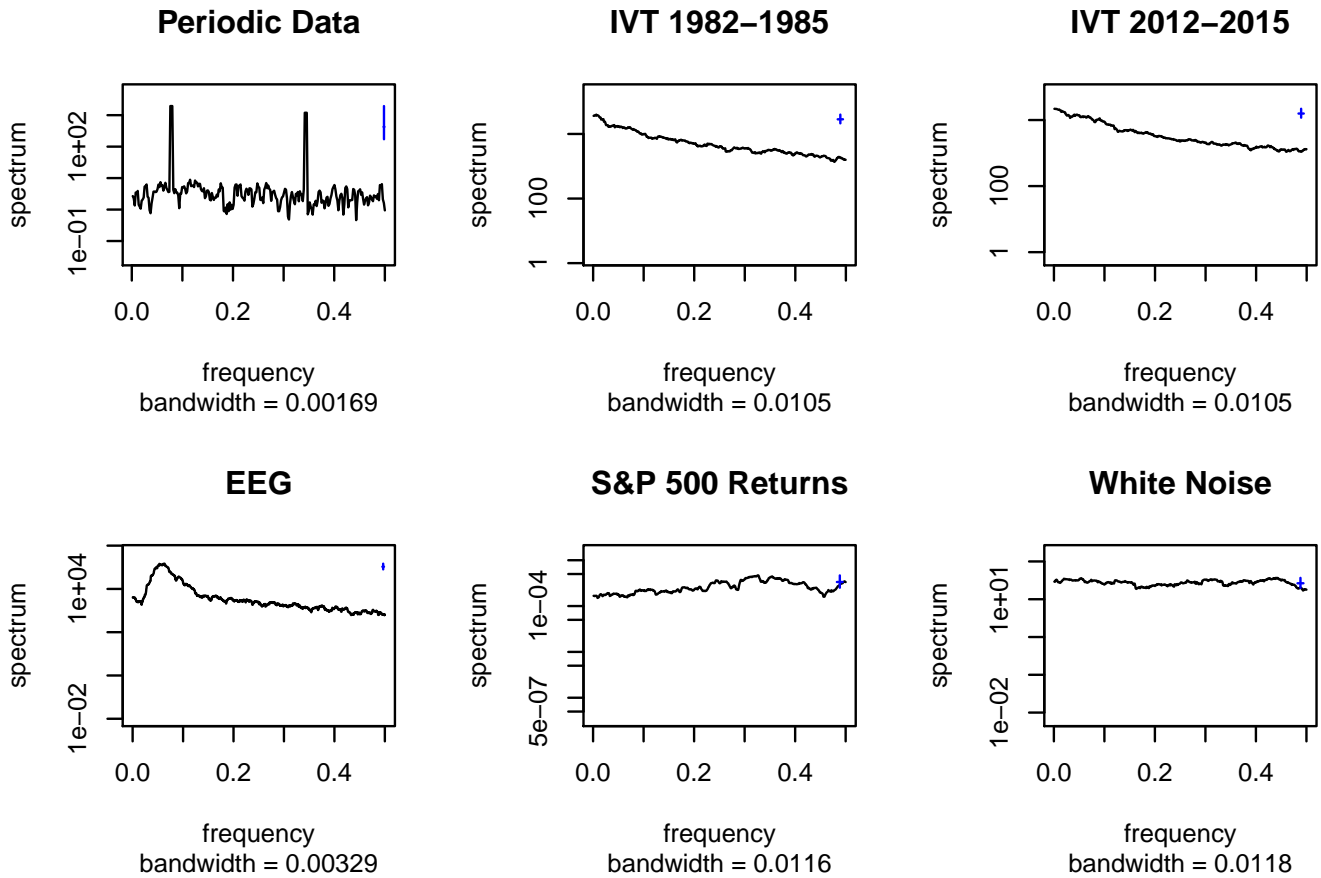
Figure 4: Smoothed Periodograms

An important observation to make is that smoothing the periodogram is a completely non-parametric way of estimating the periodogram. There exist parametric approaches for estimation as well. One approach for parametric spectrum approximation involves first assuming that a given time series process can be approximated by an autoregressive process. Then, autoregressive models of different orders are fit to the time series and information criteria are used to select the best model order. Figure 5 displays the results of applying this approach to each dataset considered in this project, where black represents the fitted spectrum for an autoregressive model with order selected by Akaike Information Criterion (AIC), and red represents the same for order selected by Bayesian Information Criterion (BIC). In Table 1 we have the orders selected by AIC and BIC for Figure 5. We note the similar shape between these parametric fits and the smoothed periodograms in Figure 4.
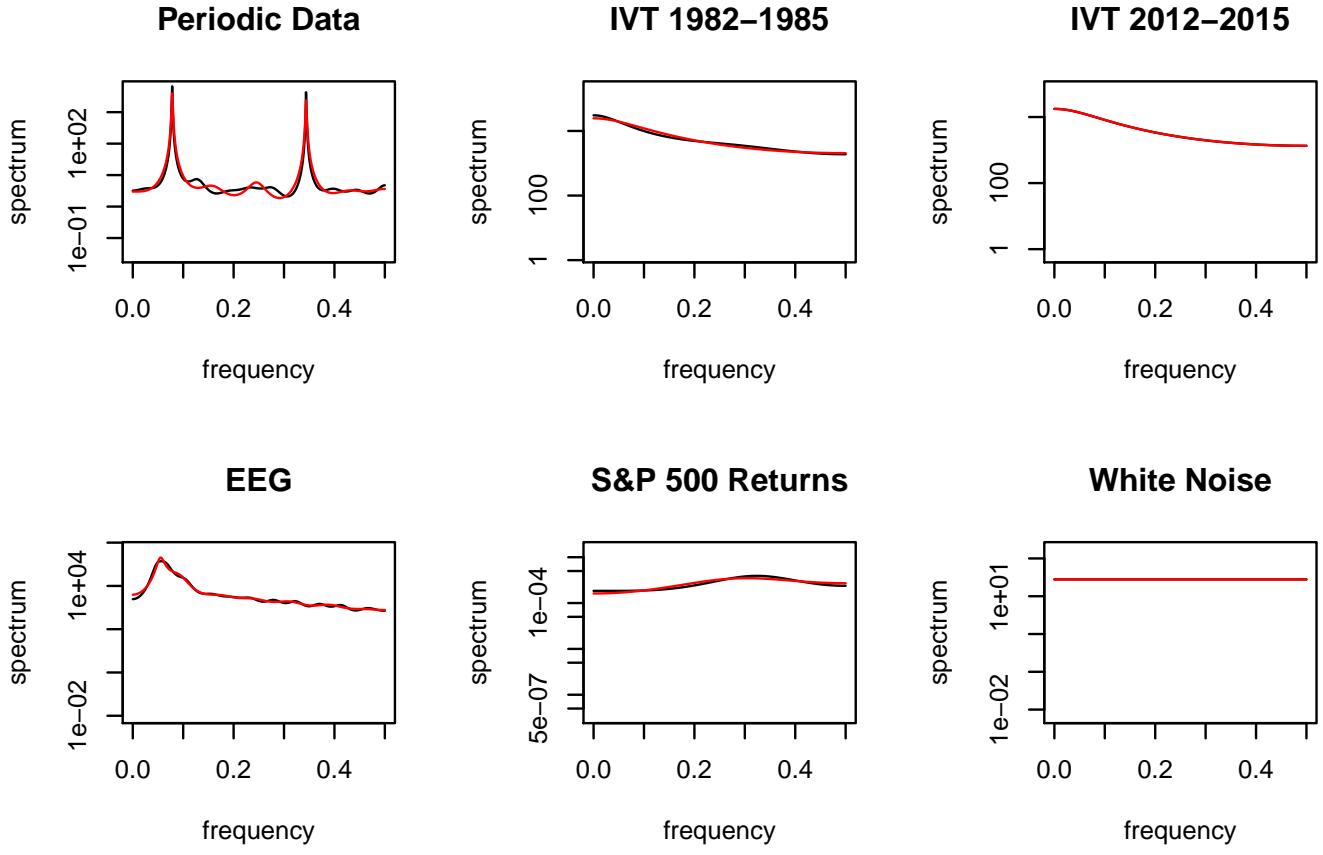
Figure 5: Parametric Spectrum Estimates

Dataset	Order Selected by AIC	Order Selected by BIC
Periodic Data	19	12
EEG	24	14
IVT 1982-1985	3	1
S&P 500 Returns	3	2
IVT 2012-2015	1	1
White Noise	0	0

Table 1: Model Order Selected by AIC and BIC

Next, we explore Bayesian approaches for spectral density estimation. Each of these approaches relies on Whittle's likelihood approximation, which is an asymptotic result regarding the likelihood of a realization of a stationary Gaussian process with spectral density $f(\omega)$. $I(\omega_j)$ is again a periodogram value. Here, we define $g(x_1, ..., x_n)$ as the joint distribution of n realizations of a stationary Gaussian process. Then the negative log likelihood can be approximated as follows (Rosen et al., 2012):

$$-log\ g(x_1, ..., x_n) \simeq \sum_{j=0}^{n-1} log\ f(\omega_j) + \frac{I(\omega_j)}{f(\omega_j)}.$$

11

Using this likelihood, we can implement Bayesian approaches for spectral density estimation. Rosen, Wood and Stoffer (2012) specify one such approach, and name it Adaptspec. The Adaptspec approach works with the log spectral density $g(\omega_j)$, and the log periodogram, $y(\omega_j)$, using the log linear model, $y(\omega_j) = g(\omega_j) + \epsilon_j$, where $\epsilon_j \sim log(\chi_2^2/2)$ for $j = 1, ..., n/2 - 1$ and $\epsilon_j \sim log(\chi_1^2)$ for $j = 0, n/2$. Seeking to approximate $g(\omega_j)$, Rosen et al. represent $g(\omega_j)$ as a combination of linear and nonlinear components, where $h(\omega_j)$ represents the nonlinear component.

$$g(\omega_j) = \alpha_0 + \alpha_1 \omega_j + h(\omega_j)$$

A normal prior is placed on α_0 and α_1 is fixed at zero. The nonlinear component $h(\omega_j)$ is defined as a linear combination of Demmler-Reinsch basis functions, denoted $X\beta$, where X is a matrix of those basis functions evaluated at the Fourier frequencies, and β is the corresponding vector of coefficients. A prior on the coefficients is defined to be $\beta \sim N(0, \tau^2 I)$. A uniform prior is placed on τ^2. Using Whittle's likelihood approximation, an MCMC approach is then implemented to estimate the parameters α_0, β, τ^2.

Rosen, Wood and Stoffer also published the BayesSpec package in R implementing the Adaptspec approach (Rosen et al., 2017). We use the BayesSpec package to generate the plots shown in Figure 6. About the spectral density estimate in black, we have 95 percent confidence regions in red. For each plot, 1000 MCMC iterations were run and the final 800 were stored. Additionally, for all computations, the default priors were used.
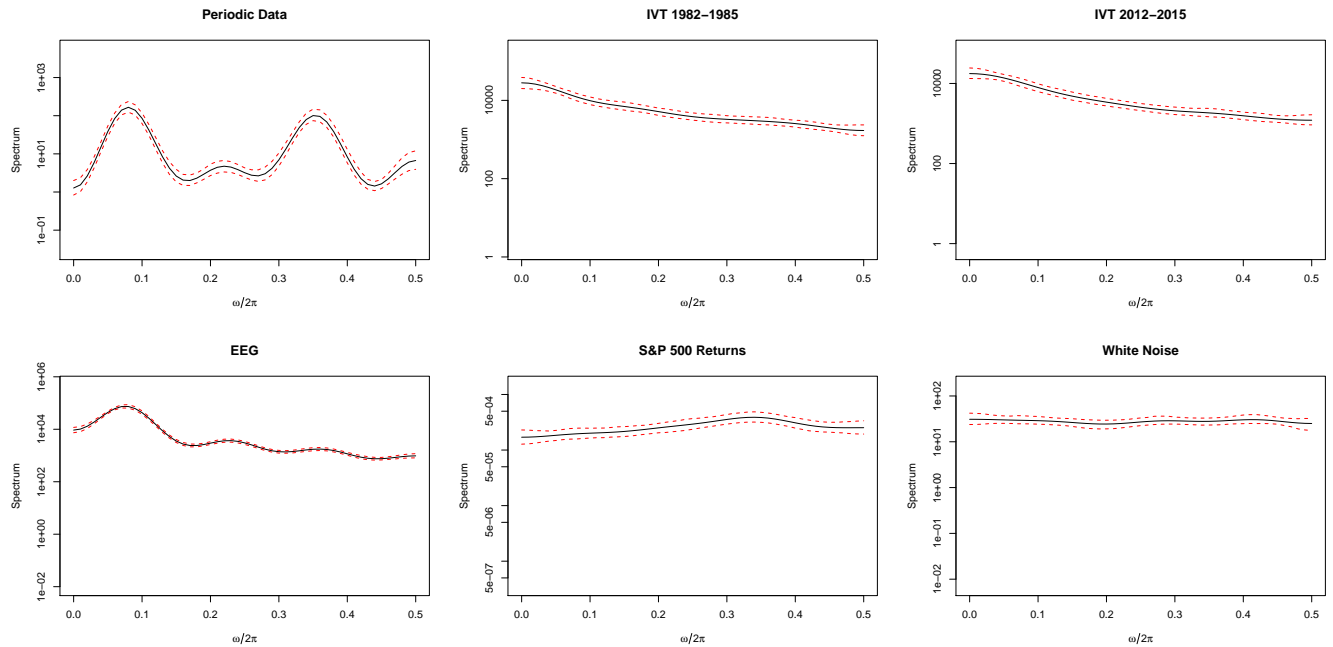

Figure 6: Bayesian Spectrum Estimates

The BayesSpec package also gives one the ability to consider non-stationary time series, by identifying segments of the time series that are locally stationary, and then computing a spectrum estimate for each locally stationary segment. Below, we outline the results of looking at the S&P 500 returns data using this approach. In Figure 7, we have a plot of the posterior likelihood on the number of segments to split the time series into. The package suggests that splitting the time series into four segments is appropriate. To split the time series into four segments, we need three break

points. For each MCMC iteration, the package provides an estimate of the index of the three break points. Figure 8 provides a summary of the distribution of where the package estimates the break points to be. Also provided is a plot of the data with the median break points overlaid. Again, the default priors were used, and inference was based on the results of 800 of the total 1000 MCMC iterations.
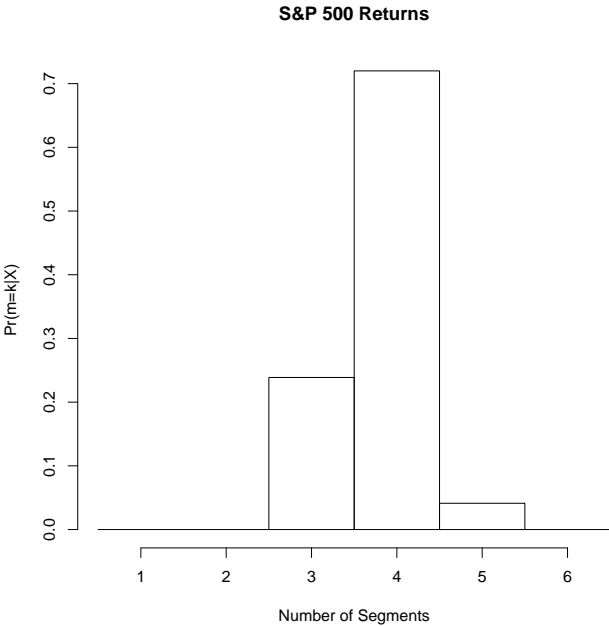
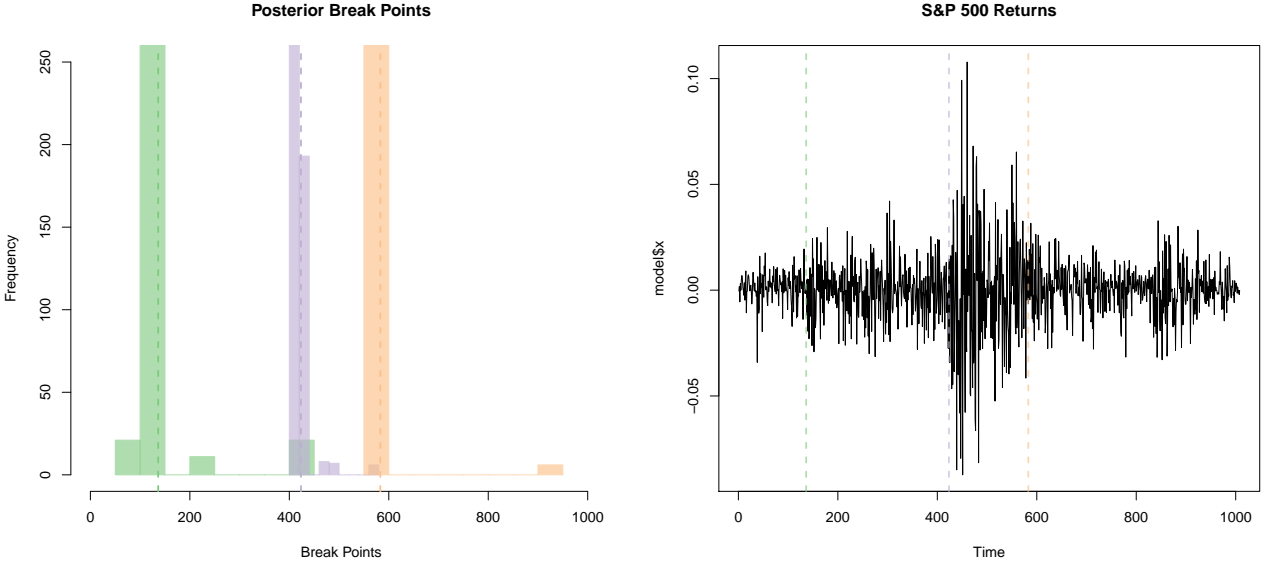

Figure 7: S&P 500 Returns: Number of Segments Selected by MCMC

Figure 8: S&P 500 Returns: Break Points

13

We see in Figure 8 that the most likely breakpoints for the series are around time points 180/1008, 410/1008 and 590/1008. Finally we provide the posterior estimates of the spectra of these four segments in Figure 9. The black line indicates the mean of posterior samples for each periodogram ordinate, and the dashed red lines indicate 95 percent confidence regions. Just as when considering the entire S&P 500 returns datset as one segment, we are unable to identify any particularly interesting behavior when splitting the time series into four segments. We conclude this by noting that the amplitude remains very small in all four segments, and also noting that in no segment is there a dramatic spike.
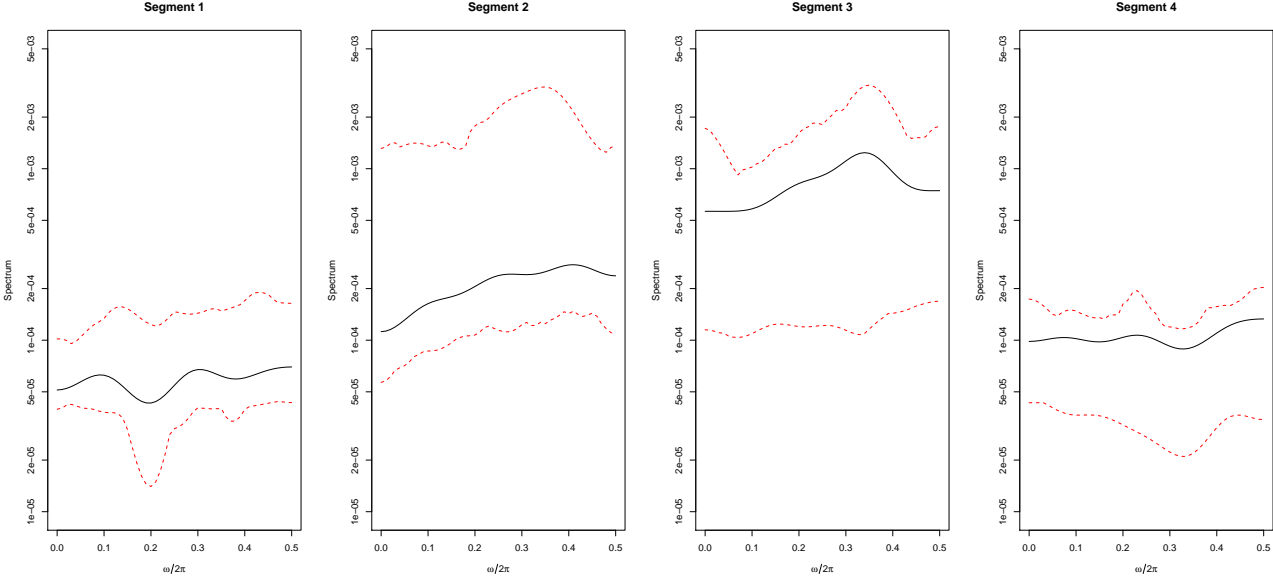

Figure 9: S&P 500 Returns: Spectrum Estimates

Next we do the same for the EEG time series. This time, the MCMC algorithm identifies five as the appropriate number of segments (Figure 10). We evaluate how the package splits the data into five segments by looking at the distribution of break points (Figure 11), and finally at the posterior estimates of the five different segments (Figure 12). Again, posterior inference is done using 800 of the total 1000 MCMC iterations and the default priors were used.
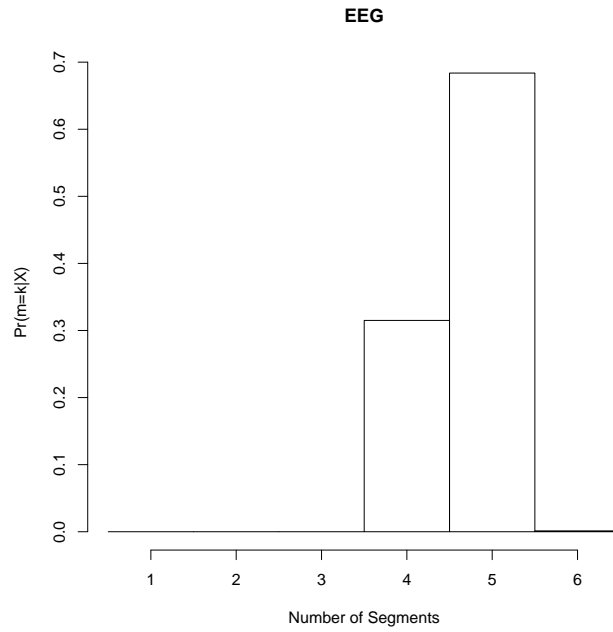
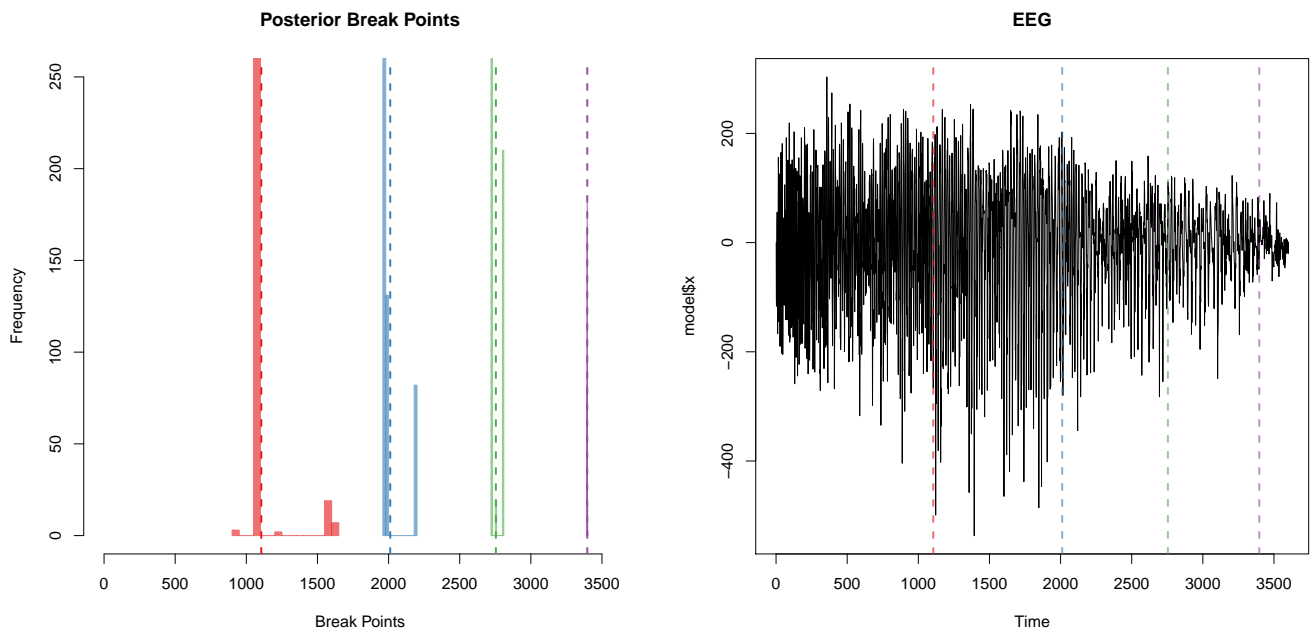Figure 10: EEG Number of Segments Selected by MCMC
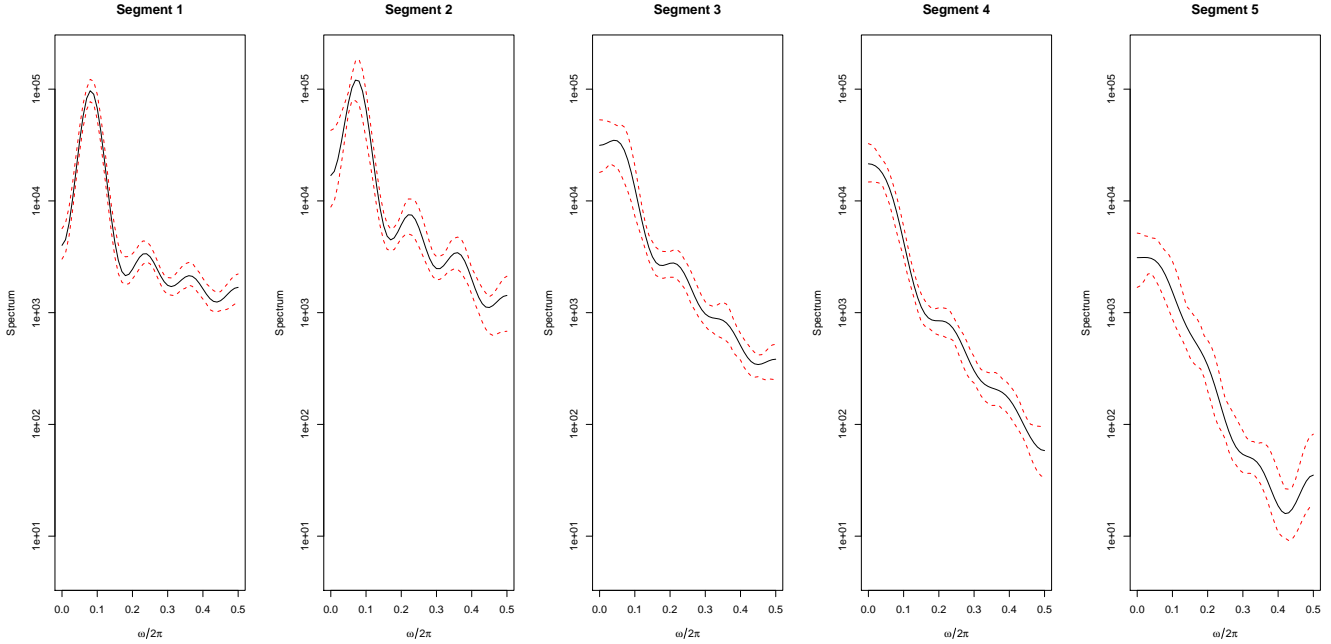

Figure 11: EEG Break Points

Figure 12: EEG Posterior Spectrum Estimates

We note that the dominant frequency of the time series decreases as we progress in time, as does the amplitude of the spectrum at the dominant frequency.

5 Quantile Spectral Analysis

Quantile spectral analysis has many similarities with traditional spectral analysis. In both cases, we can compute ACFs and create some estimates of the spectral density. The primary difference is that while traditional spectral analysis considers the mean of the data, quantile spectral analysis considers a specific quantile. For example with quantile spectral analysis, one could identify periodocity in the extremes of the data, which is not possible to identify using traditional techniques. In quantile spectral analysis, we can again define the autocorrelation as the ratio of an autocovariance at lag h to the autocovariance at lag 0. In quantile approaches, there are two nearly equivalent definitions of the autocovariance. One is the Laplace cross-covariance function (Kley, 2014), given by

$$r_{xx}(h, q_1, q_2) = Cov(I[X_t < q_1], I[X_{t+h} < q_2]), \quad q_1, q_2 \in \mathbb{R}.$$

Here q_1 and q_2 are quantiles of the strictly stationary process $\{X_t\}$. The other is the quantile cross-covariance function, (Kley, 2014) given below, where $I[\cdot]$ denotes the indicator function and F is the cumulative distribution function of any X_t:

$$r_{xx}^q(h, \tau_1, \tau_2) = Cov(I[F(X_t) < \tau_1], I[F(X_{t+h}) < \tau_2]), \quad \tau_1, \tau_2 \in [0, 1].$$

The Laplace cross-covariance function can be approximated using the following formula (Xiao, 2017), where $\hat{Q}(\tau)$ represents the τth sample quantile of the time series. This approximation is used in Figure 13 to compute the median ACF which corresponds to $\tau_1 = \tau_2 = 0.5$ for each dataset

16

considered. We expect to see similar results here as in the traditional ACF case, because the traditional ACF and the median ACF are typically similar.

$$\hat{r}_{xx}(h, \tau_1, \tau_2) = \frac{1}{n} \sum_{t=1}^{n-h} \left(I[X_t < \hat{Q}(\tau_1)] - \tau_1 \right) \left(I[X_{t+h} < \hat{Q}(\tau_2)] - \tau_2 \right).$$
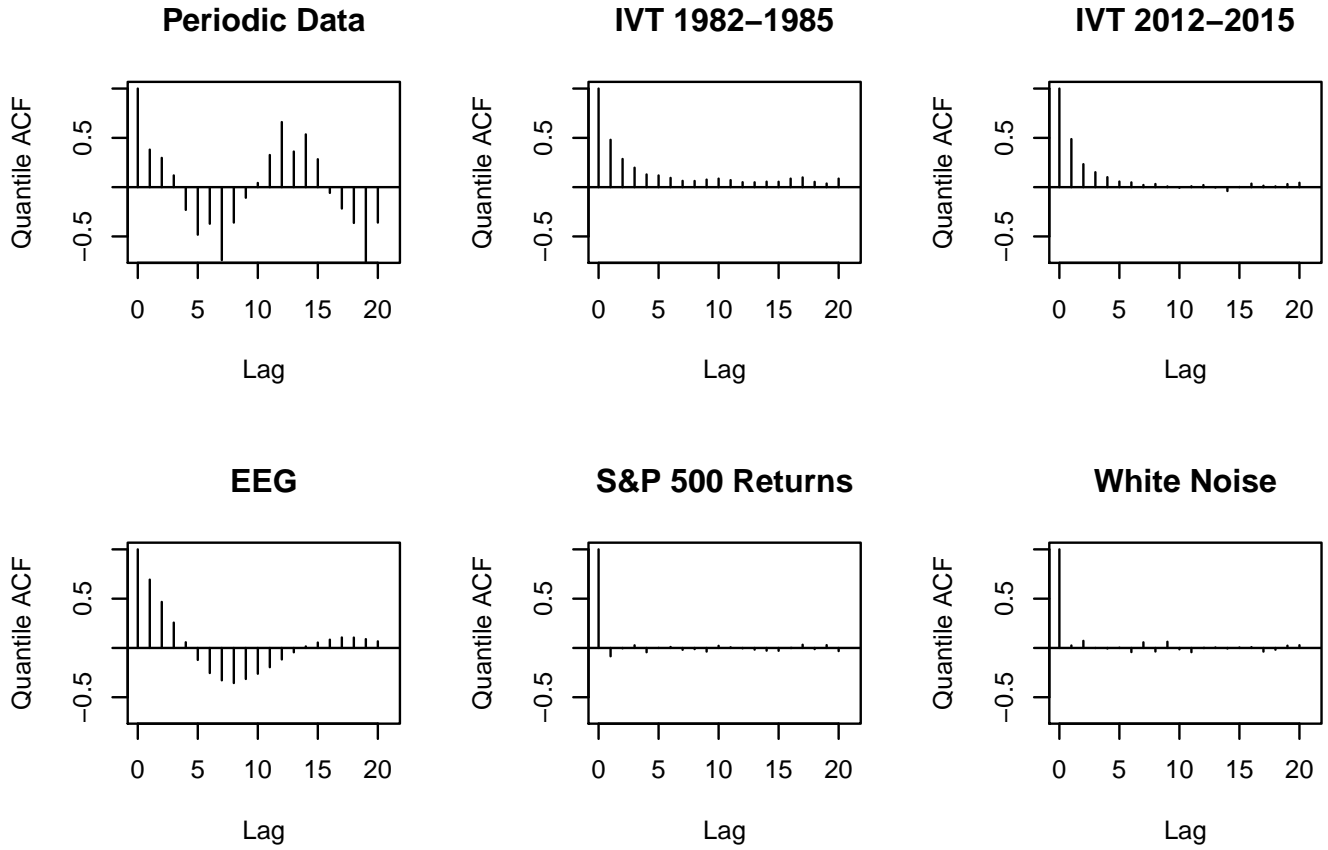

Figure 13: Quantile ACFs at Median

Comparing these results with the traditional ACFs shown in Figure 2, we find that they are indeed similar. We see again that the S&P 500 Returns quantile ACF looks similar to the white noise quantile ACF.

In Figure 14 we compute the 90th percentile ACFs for each dataset, corresponding to $\tau_1 = \tau_2 = 0.9$ to see if we can potentially uncover some dependence in the extremes.
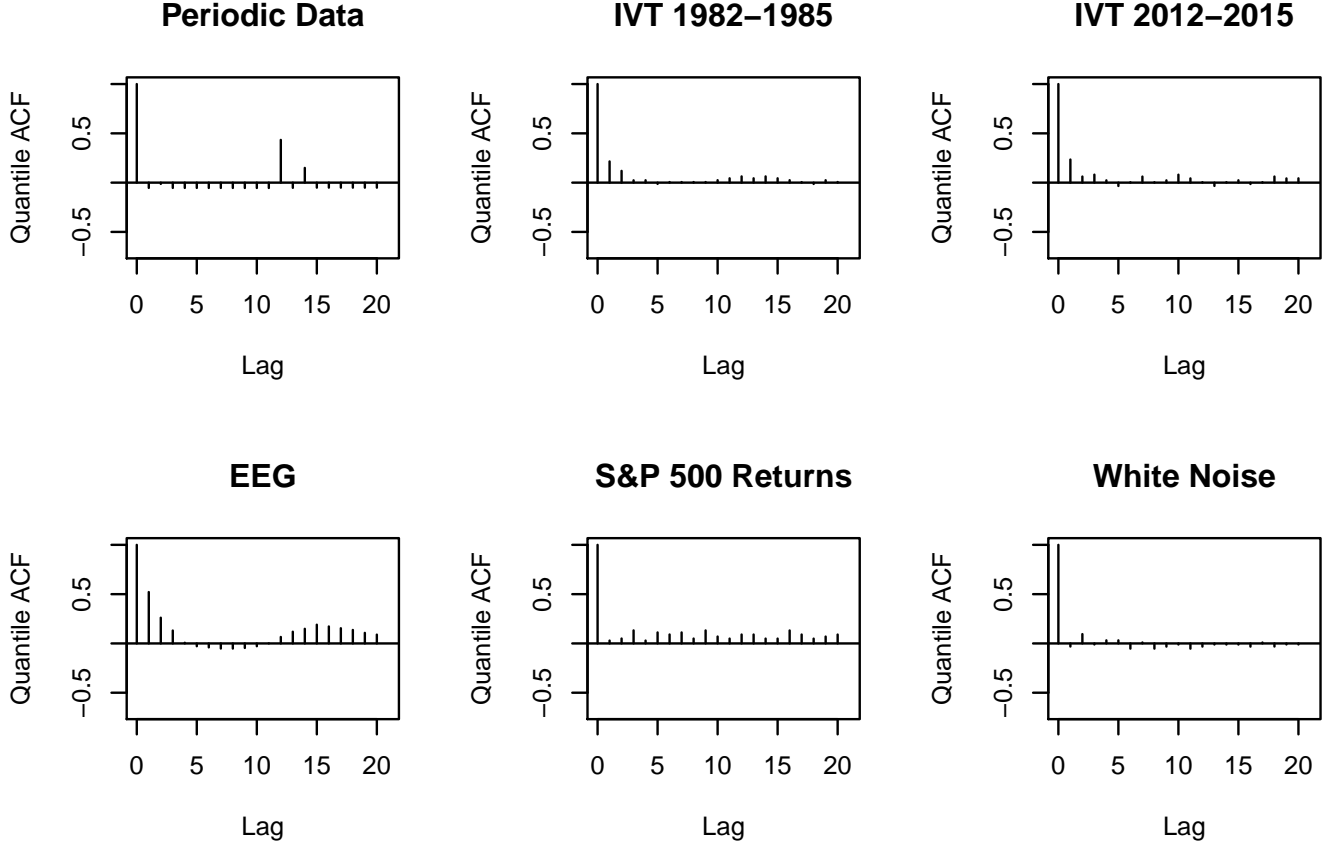
17

Figure 14: Quantile ACFs at 90th Percentile

It would seem that the overall structure remains similar for all examples but the periodic data and the S&P 500 Returns data. In the periodic data, we no longer have a roughly smooth periodic ACF, but instead something that's very small except around lags 0 and 12. In the S&P 500 data, we now have autocorrelations that are all positive and larger than before.

Just as with traditional spectral analysis, there exists a quantile version of spectral density, which has a very similar interpretation to the traditional spectral density. Specifically, the height at a given frequency indicates the contribution from that frequency in whatever quantile of the dataset being considered. One representation of the quantile spectral density, denoted S_{xx}^q, is provided below (Xiao, 2017):

$$S_{xx}^q(\omega, \tau_1, \tau_2) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} r_{xx}^q(h, \tau_1, \tau_2) e^{-ih\omega}, \quad -\pi \leq \omega \leq \pi.$$

It is possible to approximate this spectral density using methods very reminiscent of those covered in the traditional case. We could for instance estimate the quantile cross covariance and truncate at a specific lag. Alternatively, there exists a regression onto a sequence of harmonic covariates that we can perform to give us some fitted values that we can transform into a periodogram.

The primary difference between the regressions performed in the quantile and traditional cases is that in the quantile case we use check loss rather than squared error loss. The check loss function

$\rho_\tau(u)$ is defined below:

$$\rho_\tau(u) = u(\tau - I[u \leq 0])$$
$$Q_Y(\tau) = \underset{u}{\text{argmin}}\ E[\rho_\tau(Y - u)].$$

We can use the check loss function to estimate coefficients in a harmonic regression model. We estimate those coefficients in the following way (Kley et al., 2014):

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

Here, $\boldsymbol{\beta}_\tau(j/n)$ represents the vector of $\beta_{1,\tau}$ and $\beta_{2,\tau}$ corresponding to frequency $\frac{2\pi j}{n}$. Again, if a periodic signal at the particular frequency we are considering is not contributing in the τth quantile of our data, we would expect β_1 and β_2 to not be particularly large. If, however, a signal is present, then we would expect at least one of β_1 and β_2 to be large. Just as in the traditional case, we can define a quantile periodogram as a function of these two coefficients. That function is defined below (Kley et al., 2014):

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

We note that the quantile periodogram is guaranteed to be real if $\tau_1 = \tau_2 = \tau$, and that when $\tau_1 = \tau_2$, the equation for the raw periodogram reduces down to $\frac{n}{4}(\beta_1^2 + \beta_2^2)$, which is exactly the same form as in the traditional spectral analysis case.

The quantile periodogram has a very similar interpretation to the traditional periodogram, in that higher values at a particular frequency indicate greater contribution of that frequency in the data. Also, much as with the traditional approach, this regression approach to computing periodogram ordinates is computationally inefficient. Again a more computationally efficient way to compute these quantities based on the Discrete Fourier Transform exists. This approach is notably different from the traditional case in that now the DFT approach and the regression approach do not yield identical estimates of periodogram values. The two periodograms will be asymptotically equivalent up to a scaling constant, but not identical in practice.

We can write the quantile periodogram, denoted L_{xx}, as a transformation of the sample quantile cross-covariance function. That can be rewritten as the product of two DFTs, where instead of taking the transform of the series directly, we transform the clipped series, in which x_t is sent to 1 if it is below the τth sample quantile and to 0 otherwise. (Kley et al., 2014)

$$L_{xx}(\omega, \tau_1, \tau_2) = \frac{1}{2\pi} \sum_{h=-(n-1)}^{n-1} \hat{r}_{xx}(h, \tau_1, \tau_2)e^{-i\omega h}$$
$$= \frac{1}{2\pi}d_{\tau_1}(-\omega)d_{\tau_2}(\omega)$$
$$\text{where} \quad d_\tau(\omega) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} I[X_t \leq \hat{Q}(\tau)]e^{-i\omega t}.$$

Here, ω will take the usual $2\pi j/n$ form. In Figure 15 we consider the results of the regression approach, when $\tau_1 = \tau_2 = 0.5$.
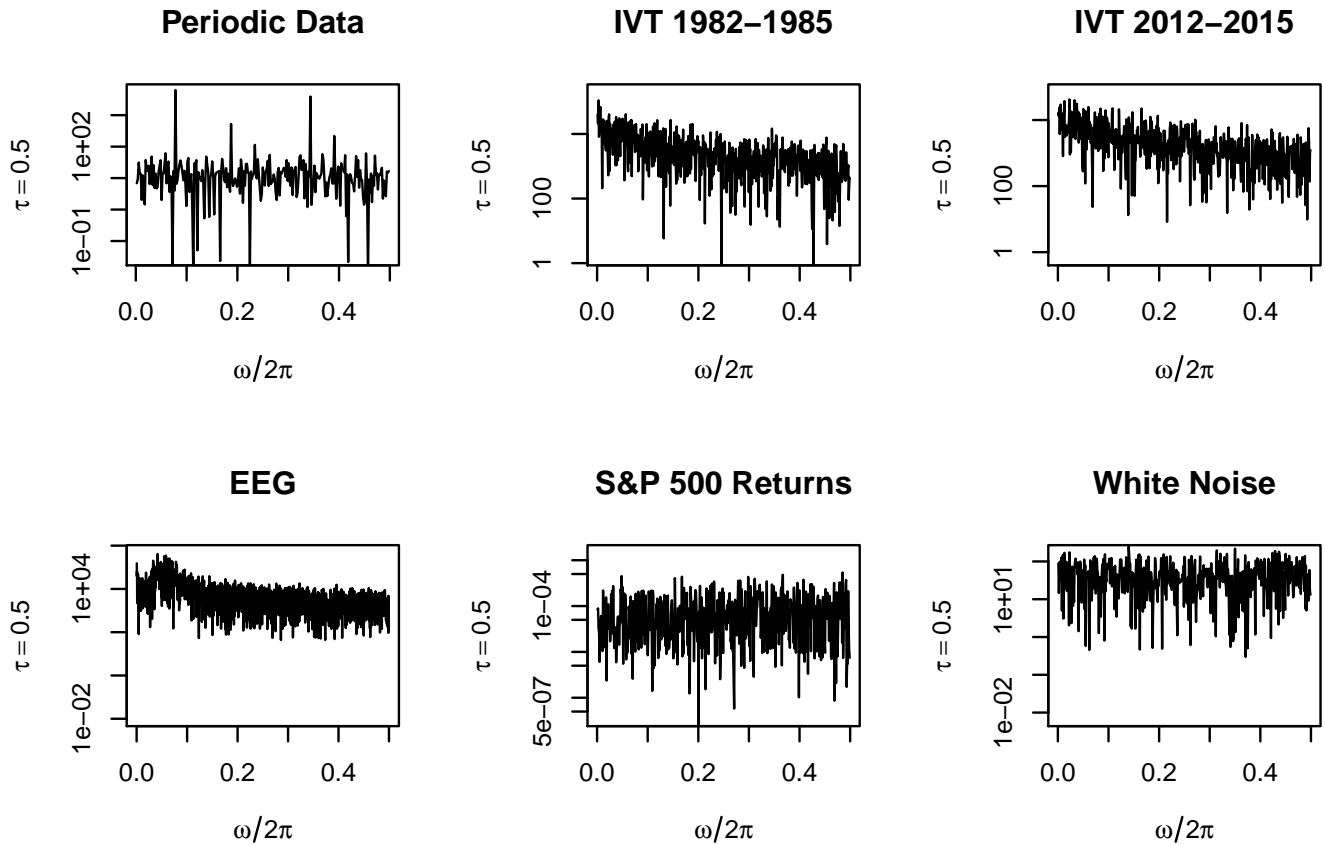
Figure 15: Quantile Periodogram at Median Computed Using Regression Approach

In Figure 16 we plot the periodograms yielded by the DFT approach when $\tau_1 = \tau_2 = 0.5$. The DFT approach to computing periodograms is substantially faster than the harmonic regression approach, taking about a tenth of the time. We note that the two approaches do not yield identical quantile periodograms.
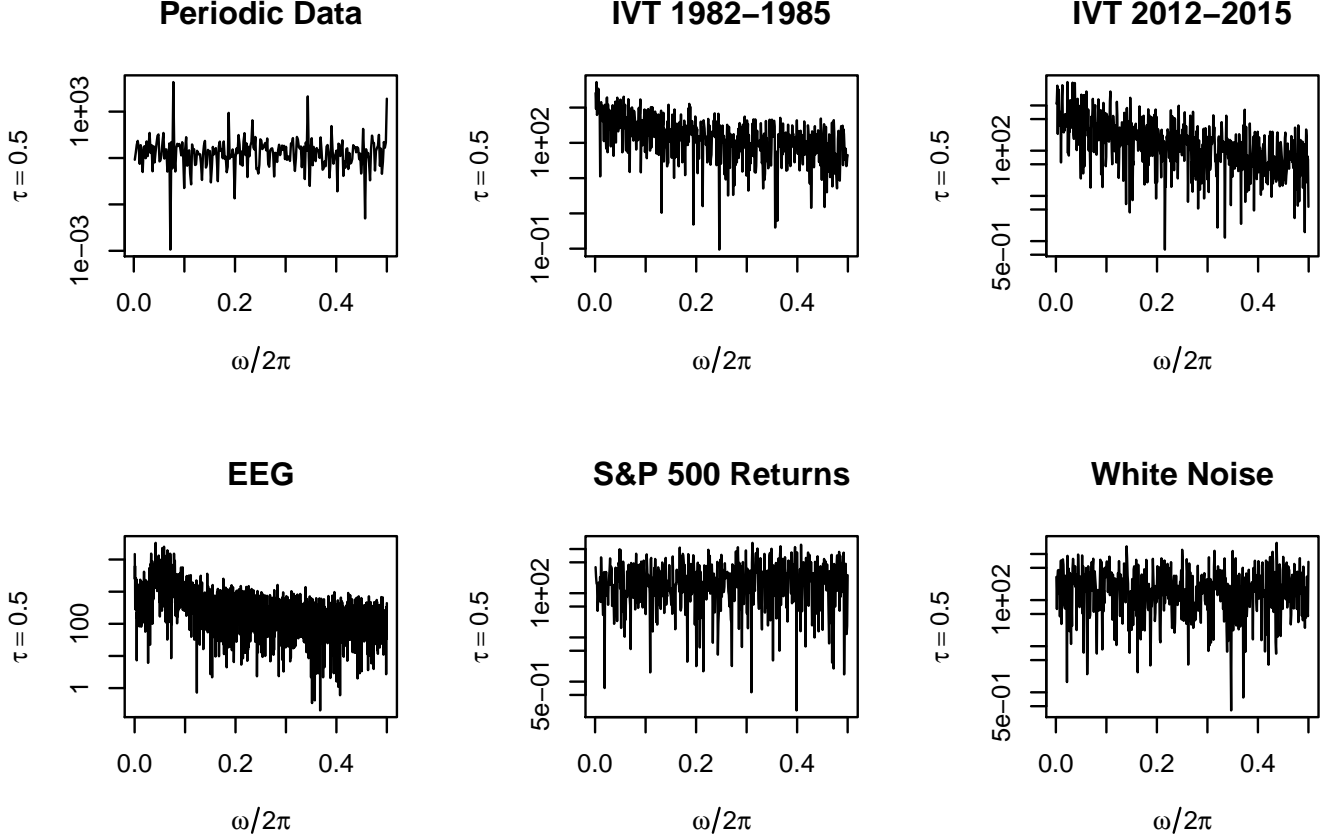
Figure 16: Quantile Periodogram at Median Computed Using DFT Approach

Just as in Figure 15, in Figure 16 we observe very similar periodograms for the S&P 500 Returns data and the white noise data. With both computation approaches, the overall structure is very similar, but the exact shape and scaling are not identical. Under the assumption of strict stationarity (Dette et al., 2015), the scaling can be rectified by adjusting according to the following equation that incorporates the asymptotic error. Here f indicates the density of X_t, and F the CDF of X_t (Xiao, 2017). The $o_p(1)$ indicates that the difference between the estimates of the spectral density converges in probability to zero as the number of observations goes to infinity.

$$\frac{n}{4}\beta_{\tau_1}(j/n)^T \begin{bmatrix} 1 & i \\ -i & 1 \end{bmatrix} \beta_{\tau_2}(j/n) = \frac{1}{f(F^{-1}(\tau_1))f(F^{-1}(\tau_2))} \sum_{h=-(n-1)}^{n-1} \hat{r}_{xx}^q(h, \tau_1, \tau_2)e^{-ih\omega_j} + o_p(1).$$

Additionally, in Figure 17 we provide 90th quantile periodograms for the datasets considered to demonstrate some of the unique capabilities of this approach. In particular, we note that the S&P 500 Returns data and white noise data no longer have similar looking periodograms.
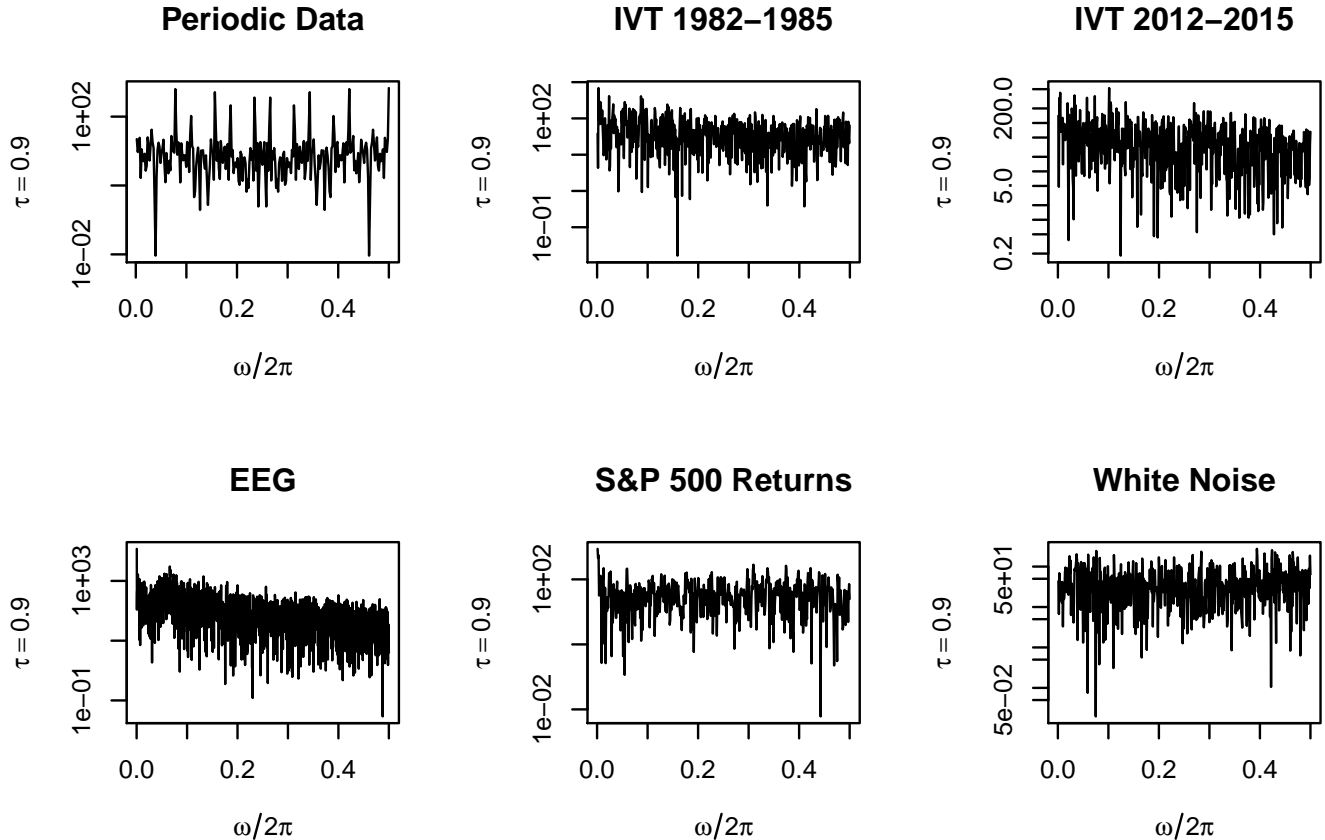
Figure 17: Quantile Periodogram at 90th Percentile Computed Using DFT Approach

Each of the above periodograms is very noisy. This can be expected because, much like the traditional periodogram, the quantile periodogram is not a consistent estimator of the quantile spectral density. Similar approaches to achieving a consistent estimator exist in both cases. The quantspec package (Kley, 2014) employs an identical approach to achieve a consistent estimator as was covered in the traditional periodogram section. Namely, the package smooths the raw quantile periodogram. We denote the value of the smoothed quantile periodogram at a particular frequency U and define U below:

$$U(\tau_1, \tau_2, \omega) = \frac{2\pi}{n} \sum_{s=1}^{n-1} W_n(\omega - 2\pi s/n) L_{xx}(2\pi s/n).$$

Here W_n is a sequence of weights that sum to 1, and can correspond to as large or small a window as one feels is appropriate. In Figure 18 we include smoothed periodograms for the datasets considered when $\tau_1 = \tau_2 = 0.5$. Regarding the degree of averaging, the quantspec package defaults to a bandwidth of 0.1 using the Daniell kernel.
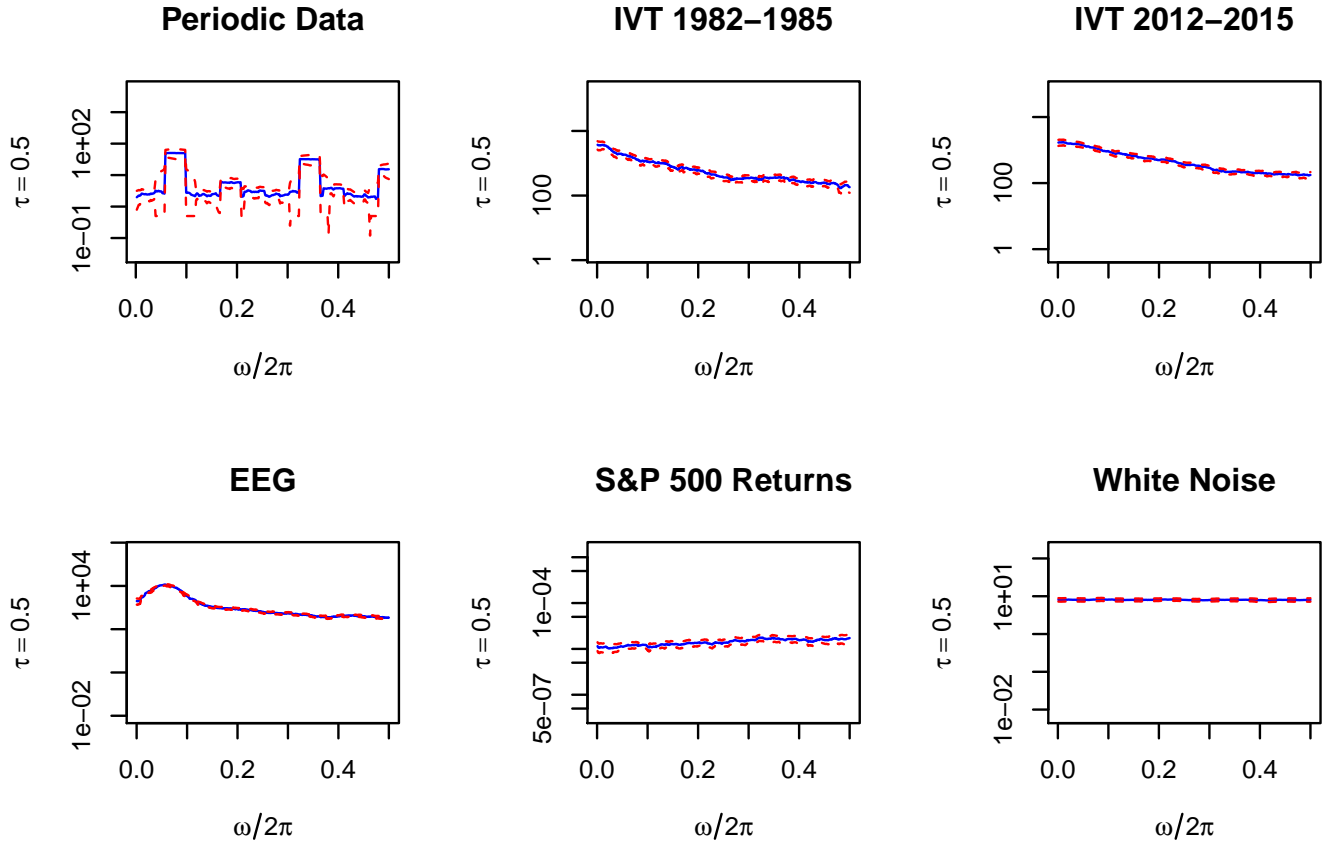
Figure 18: Smoothed Quantile Periodgram at Median

    The dotted red lines denote a 95 percent confidence interval based on the asymptotic normality of the smoothed quantile periodogram ordinates. While this is the default method, the quantspec package can also produce confidence intervals on the smoothed periodogram using a bootstrapping procedure.

    As a final result of this project, we modify the BayesSpec package's function for estimating spectrums to utilize quantile periodograms, rather than traditional periodograms. This allows us to create a Bayesian approximation of the quantile periodogram. We do so for each dataset at the median and at the 90th percentile and the results are in Figure 19 and Figure 20, respectively. Each figure is computed using 800 of the total 1000 MCMC iterations. The black line indicates the mean of posterior samples for each periodogram ordinate, and the dashed red lines indicates 95 percent confidence regions.
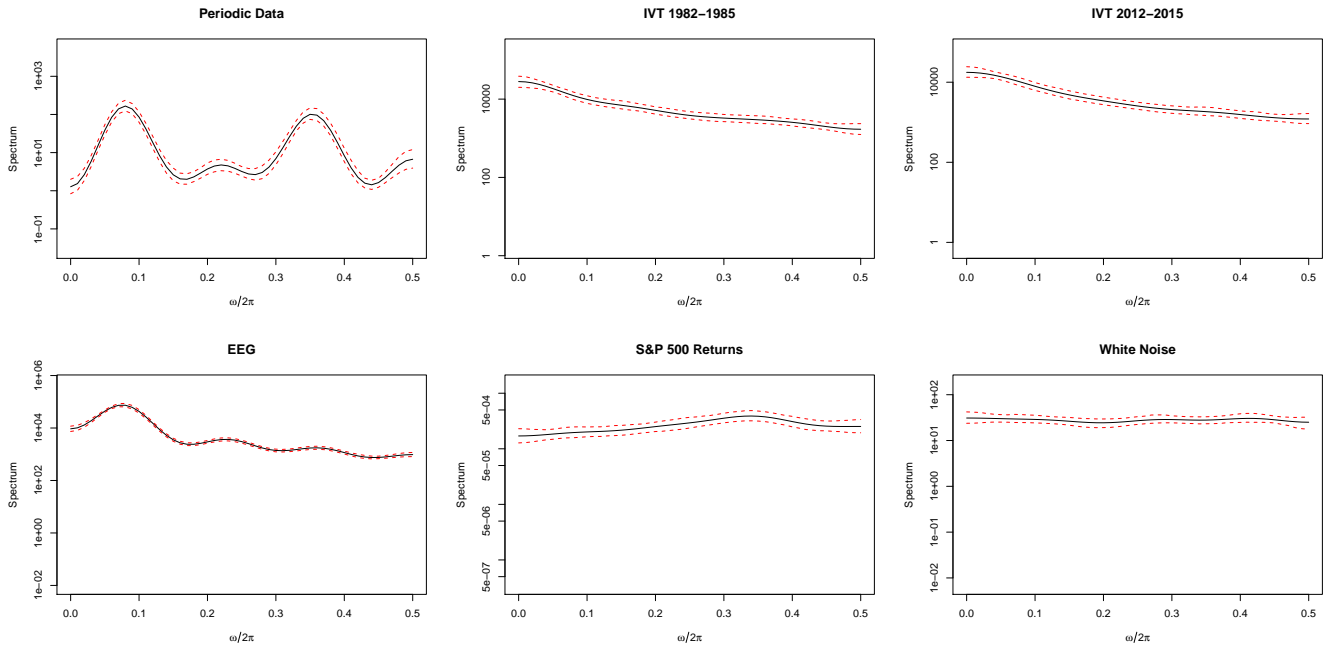
Figure 19: Bayesian Quantile Spectrum Estimates for Median

Comparing Figures 6 and 19, we note that swapping the traditional periodogram for the quantile periodogram at the median makes little difference in the scale and shape of the resulting estimates of the spectral density. In particular, we again see no dramatic difference in shape between the white noise plot and the S&P 500 Returns plot. The similarity between the white noise spectrum and S&P 500 Returns spectrum is not maintained when we consider the 90th percentile in Figure 20. In the 90th percentile, we observe a spike at low frequencies in the S&P 500 Returns spectrum, which is consistent with earlier results. In contrast, the white noise plot continues to be relatively flat.
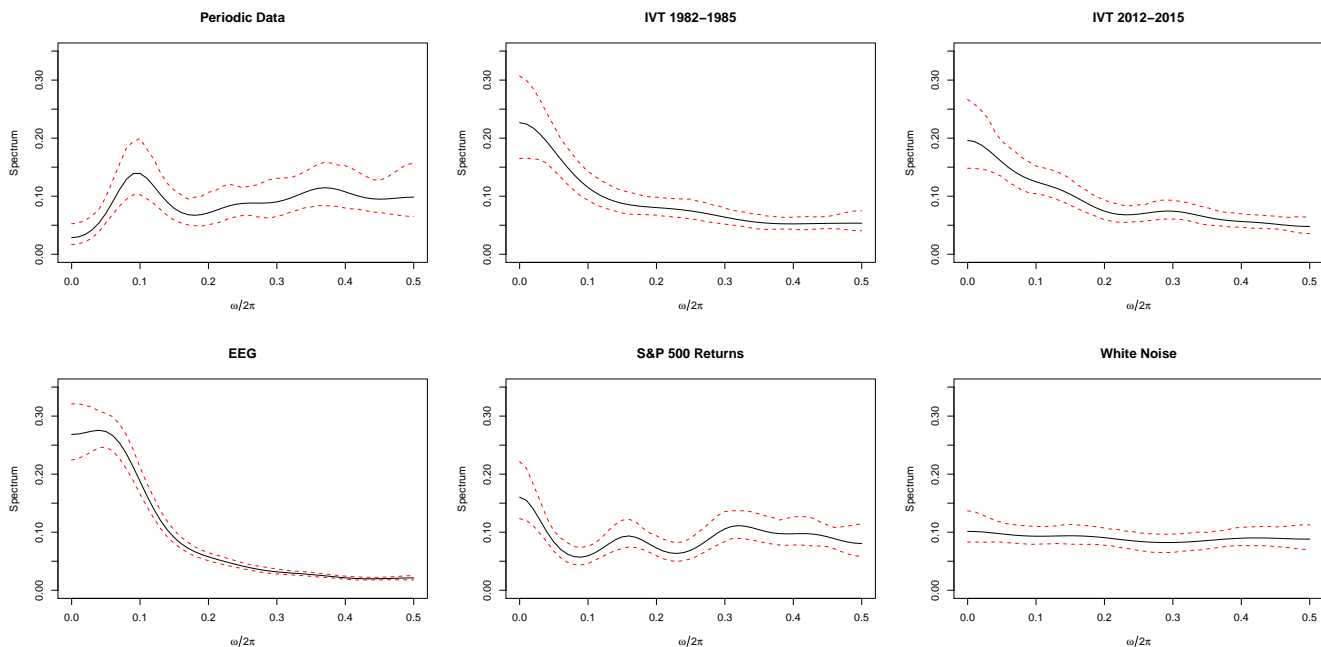
Figure 20: Bayesian Quantile Spectrum Estimates for 90th Percentile

6   Conclusion

In this project, we compared traditional spectral analysis to quantile spectral analysis. In particular, we introduced traditional and quantile versions of the periodogram and ACF, and then discussed a number of methods for computing an estimate of the spectral density based on these quantites. We highlighted the similarities between the traditional periodogram and the quantile periodogram at the median, but noted that the traditional spectral density is unable to fully capture some characteristics of a time series. This deficiency was exemplified most clearly in the comparison of spectral density estimates of the white noise time series and the S&P 500 returns time series. Both time series produced nearly identical ACFs and periodograms, but are visually very different. In contrast, we were able to characterize the non-stationary variance in the S&P 500 returns data using quantile techniques. We concluded the project with a Bayesian approach to estimating a quantile spectral density, but there is more work to be done in this area. A natural extension of the work done in this project would be to define a quantile form of Whittle's likelihood and to explore the theoretical properties of periodograms resulting from this quantity.

References

Dette, H., M. Hallin, T. Kley, and S. Volgushev (2015, 05). Of copulas, quantiles, ranks and spectra: An l_1-approach to spectral analysis. *Bernoulli 21*(2), 781–831.

Guan, B. and D. E. Waliser (2015). Detection of atmospheric rivers: Evaluation and application of an algorithm for global studies. *Journal of Geophysical Research: Atmospheres 120*(24), 12514–12535.

Kley, T. (2014). Quantile-based spectral analysis in an object-oriented framework and a reference implementation in R: The quantspec package. *arXiv e-prints*, arXiv:1408.6755.

Kley, T., S. Volgushev, H. Dette, and M. Hallin (2014). Quantile spectral processes: Asymptotic analysis and inference. *arXiv e-prints*, arXiv:1401.8104.

Rosen, O., S. Wood, and D. S. Stoffer (2012). Adaptspec: Adaptive spectral estimation for nonstationary time series. *Journal of the American Statistical Association 107*(500), 1575–1589.

Rosen, O., S. Wood, and D. S. Stoffer (2017). Package 'bayesspec'.

Shumway, R. H. and D. S. Stoffer (2017). *Time Series Analysis and Its Applications: With R Examples*. New York: Springer.

West, M., R. Prado, and A. D. Krystal (1999). Evaluation and comparison of EEG traces: Latent structure in nonstationary time series. *Journal of the American Statistical Association 94*(448), 1083–1095.

Xiao, Z. (2017). *QAR and Quantile Time Series Analysis*. Chapman and Hall/CRC.