

14.384 Time Series Analysis, Fall 2007
 Professor Anna Mikusheva
 Paul Schrimpf, scribe
 September 18, 2007
 revised September 21, 2009

Lecture 4

Spectrum

Review

Recall the spectrum is

$$S(\omega) = \sum_{j=-\infty}^{\infty} e^{-i\omega j} \gamma_j$$

note that $\gamma_j = \gamma_{-j}$, so $S(\omega)$ is real valued and

$$S(\omega) = \gamma_0 + 2 \sum_{j=1}^{\infty} \gamma_j \cos(j\omega).$$

The last equation implies that S is a symmetric function: $S(\omega) = S(-\omega)$, and periodic $S(\omega) = S(\omega + 2\pi)$. That is, it is enough to depict S on the interval $[0, \pi]$. One can also prove that since γ_j are autocovariances we will have $S(\omega) \geq 0$ for all ω .

Recall that we can recover the covariances from the spectrum using the inverse Fourier transform

$$\gamma_j = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i\omega j} S(\omega) d\omega$$

The ratio $\frac{S(\omega)}{2\pi}$ is often called *spectral density*, because the “cdf” $F(\omega) = \int_{-\pi}^{\omega} \frac{S(\lambda)}{2\pi} d\lambda$ defined out of it has the following property:

$$\begin{aligned} \gamma_0 &= \int_{-\pi}^{\pi} dF(\omega) \\ \gamma_j &= \int_{-\pi}^{\pi} e^{i\omega j} dF(\omega) \end{aligned}$$

Cramer’s Representation

The spectrum will lead us to a new way of representing a time series. We will sketch an argument to show this.

Let us consider two random variables A and B such that $EA = EB = 0$, $EA^2 = EB^2 = \sigma^2$, and $EAB = 0$. Let us define a complex-valued random variable $Z = A + iB = Re^{i\phi}$ with $ER^2 = EA^2 + EB^2 = 2\sigma^2$. Now let us consider a time series

$$x_t = \frac{Ze^{i\lambda t} + \bar{Z}e^{-i\lambda t}}{2} = \frac{Re^{i(\lambda t + \phi)} + Re^{-i(\lambda t + \phi)}}{2} = R \cos(t\lambda + \phi).$$

The process x_t is a sinusoid with the random amplitude R , period $\frac{2\pi}{\lambda}$ and a random phase ϕ . Another way of writing x_t is

$$x_t = A \cos(\lambda t) - B \sin(\lambda t)$$

We can notice that x_t is a weakly stationary process with the covariance structure

$$\begin{aligned}\gamma_0 &= E x_t^2 = E(A \cos(\lambda t) - B \sin(\lambda t))^2 = \sigma^2 \\ \gamma_k &= E x_t x_{t+k} = E[A \cos(\lambda t) - B \sin(\lambda t)](A \cos(\lambda(t+k)) - B \sin(\lambda(t+k))) = \\ &= \sigma^2(\cos(\lambda(t+k)) \cos(\lambda t) + \sin(\lambda(t+k)) \sin(\lambda t)) = \sigma^2 \cos(\lambda k) = \sigma^2 \frac{(e^{i\lambda k} + e^{-i\lambda k})}{2}\end{aligned}$$

In the first line we used $\cos^2(x) + \sin^2(x) = 1$, and in the third : $\cos(x+y) = \cos(x)\cos(y) - \sin(x)\sin(y)$.

Now let's compare the last to the formula:

$$\gamma_j = \int_{-\pi}^{\pi} e^{i\omega j} dF(\omega)$$

We can see that F has two mass points: it puts weight $\frac{\sigma^2}{2}$ at λ and $-\lambda$.

Let $\lambda_j \in [0, \pi]$ be evenly spaced fixed points with $j = 1, \dots, n$, and consider a process

$$\begin{pmatrix} A(\lambda_j) \\ B(\lambda_j) \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_j^2 & 0 \\ 0 & \sigma_j^2 \end{pmatrix} \right).$$

Assume that A and B at different λ_j 's are independent. Consider $Z(\lambda_j) = A(\lambda_j) + iB(\lambda_j) = R_j e^{i\phi_j}$. We can show the following:

1. $E Z(\lambda_j) = 0$
2. $\text{Var}(Z(\lambda_j)) = E Z(\lambda_j) \overline{Z(\lambda_j)} = E R_j^2 = \sigma_j^2$
3. $E Z(\lambda_j) \overline{Z(\lambda_k)} = 0$ for $j \neq k$

Remark 1. I want to remind you that for two complex valued random variables ξ and η the covariance is defined as $\text{cov}(\xi, \eta) = E[(\xi - E\xi)(\overline{\eta - E\eta})]$.

Remark 2. Z is a discrete orthogonal process.

Suppose

$$\begin{aligned}x_t &= \sum_{j=1}^n \frac{Z(\lambda_j) e^{it\lambda_j} + \overline{Z(\lambda_j)} e^{-it\lambda_j}}{2} \\ &= \sum_{j=1}^n \cos(\lambda_j t) A(\lambda_j) - \sin(\lambda_j t) B(\lambda_j)\end{aligned}$$

Also

$$\begin{aligned}\gamma_k &= E[x_{t+k} x_t] = \sum_{j=1}^n \sigma_j^2 \frac{e^{i\lambda_j k} + e^{-i\lambda_j k}}{2} \\ &= \int_{-\pi}^{\pi} e^{ik\omega} d\tilde{F}(\omega)\end{aligned}$$

where $\tilde{F}(\omega)$ puts weight $\sigma_j/2$ at points $-\lambda_j$ and λ_j .

Let's double the number of points naming them ω_j by considering λ_j and their negatives. That is $\{\omega_j\} = \{\lambda_j\} \cup \{-\lambda_j\}$. Let $\tilde{F}(\omega) = \sum_{\omega_j < \omega} \frac{\sigma_j^2}{2}$ is a the cdf of a discrete process. Assume that we define $Z(\omega_j)$ for negative argument as $Z(\omega_j) = \overline{Z(-\omega_j)}$. Assume $\tilde{y}(\omega) = \sum_{\omega_j < \omega} Z(\omega_j)/2$, so that

$$x_t = \sum_{j=1}^n \frac{Z(\lambda_j)}{2} e^{it\lambda_j} = \int_{-\pi}^{\pi} e^{it\omega} d\tilde{y}(\omega)$$

where $d\tilde{y}(\lambda_j) = \frac{Z(\lambda_j)}{2}$, and $\text{Var}(d\tilde{y}(\lambda_j)) = d\tilde{F}(\omega)$.

Remark 3. We want to let $n \rightarrow \infty$, but then we need to use Ito integrals, which we haven't covered.

Definition 4. A mean zero *orthogonal increment process*, $y(\lambda)$, associates with each $\lambda \in [-\pi, \pi]$ a random variable $y(\lambda)$ such that $E[y(\lambda)^2]$ is finite for all λ and $E[y(\lambda_4 - \lambda_3)y(\lambda_2 - \lambda_1)] = 0$ for all $\lambda_1 \leq \lambda_2 \leq \lambda_3 \leq \lambda_4$

Remark 5. Brownian motion is an orthogonal increment process, but definitely not the only orthogonal increment process.

Theorem 6 (Cramer's representation). *Suppose x_t is stationary with zero mean and spectrum $S(\omega)$. Then \exists a right-continuous orthogonal increment process, $y(\lambda)$ (with $\lambda \in [-\pi, \pi]$) such that:*

1. $x_t = \int_{-\pi}^{\pi} e^{it\lambda} dy(\lambda)$ with probability 1

2. $E|y(\lambda) - y(-\pi)|^2 = \int_{-\pi}^{\lambda} \frac{S(\omega)}{2\pi} d\omega$

Remark 7. $dy(\lambda)$ will satisfy the continuous analogs of properties 1-4 given for $\mathcal{Z}(\lambda)$ above.

Remark 8. This means we can decompose x_t as a sum of processes at different frequencies. The integral above is a sum across all frequencies. $dy(\lambda)$ is a random weight for each frequency. We are interested in the existence of this decomposition because it means that we can sensibly talk about movements that happen at different frequencies, and try to isolate the frequencies that we're interested in. For example, to study business cycles, we might want to remove low frequencies (trends) and high frequencies (seasonality, measurement error). We do this to identify recessions. It can also be useful for gathering stylized facts and comparing them with theory. For example, we might think money is neutral in the long run, but not the short run. We could look at low and high frequency ω movements in money and GDP to look for evidence of this.

Filtering

Given the above results, we can think about how to remove frequencies from a process.

Definition 9. We apply a *linear filter*, $B(L)$ to x_t as $B(L)x_t = y_t$

Recall that $S_y(\omega) = |B(e^{i\omega})|^2 S_x(\omega)$. We can use this relationship to study how a given filter affects various frequencies.

Isolating the business cycle

Suppose $t = 1, \dots, T$ in quarters. Following the NBER, define the business cycle as the component with a period from 1.5 to 8 years (6-32 quarters). Then a period of 6 corresponds to the frequency $\lambda = \frac{2\pi}{6} = \frac{\pi}{3}$ and 32 corresponds to $\lambda = \frac{2\pi}{32} = \frac{\pi}{16}$. We want to decompose $x_t = b_t + \tau_t + e_t$, a business cycle b_t , trend τ_t , and seasonal e_t . To isolate b_t , an ideal filter would have

$$|B^*(e^{i\omega})|^2 = \begin{cases} 1 & \lambda \in [\frac{\pi}{16}, \frac{\pi}{3}] \\ 0 & \text{otherwise} \end{cases}.$$

This would select the desired range of frequencies. The problem is that there is no finite order polynomial that accomplishes this.

The ideal filter has an inverse Fourier transform, $B^*(e^{i\omega}) = \sum_{-\infty}^{\infty} \beta_j^* e^{-i\omega j}$, with infinitely many $\beta_j^* \neq 0$. We can't use this because our data is finite. One alternative is to just set $\beta_j^* = 0$ for $|j| > J$.

One of the possible conditions we might impose on a filter is that it kills zero-frequencies, $B(1) = 0$. It implies that we are removing a deterministic trend. Then even if we have non-stationary time-series, the filtered series will be stationary.

Figure 1: Baxter King (1999) Figure 2

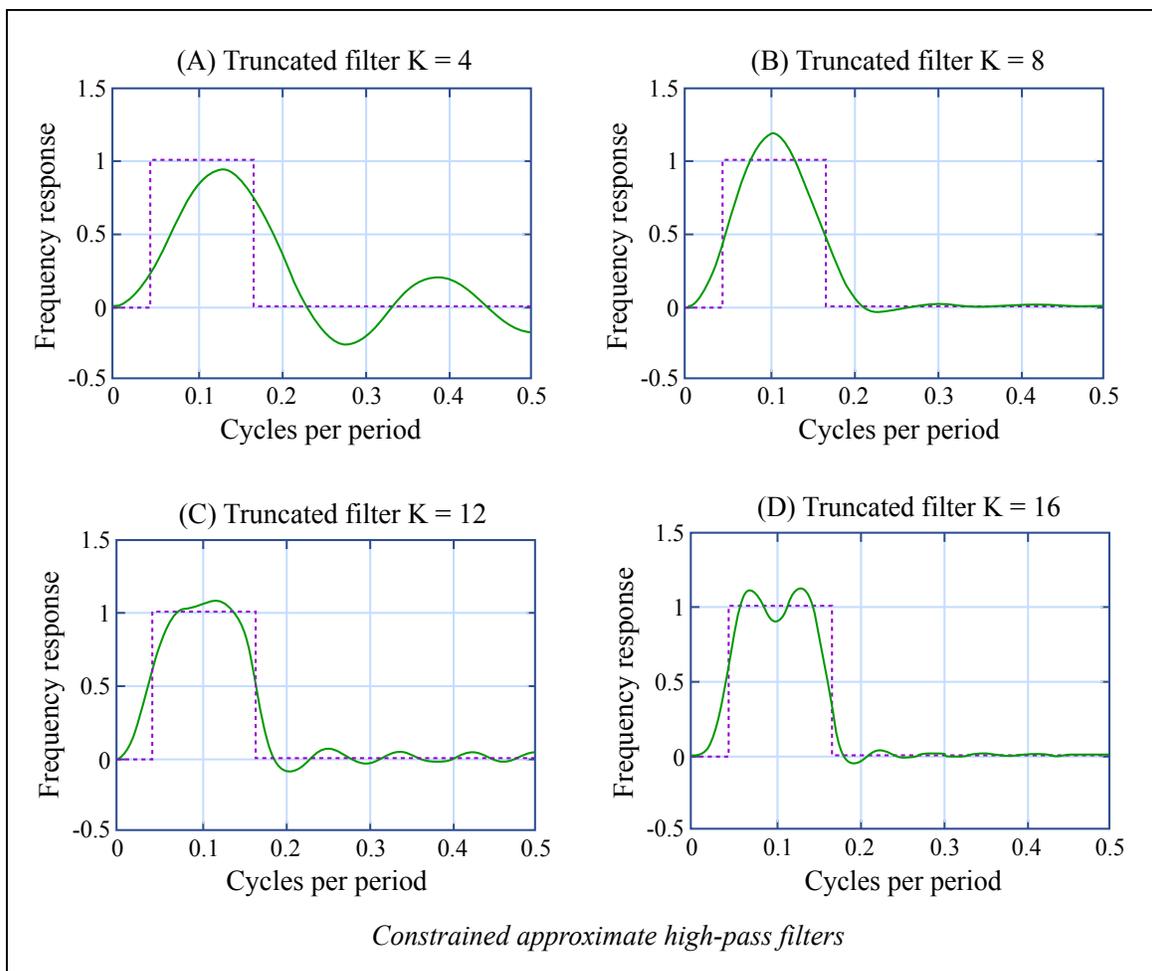


Image by MIT OpenCourseWare.

The Baxter-King approximation to the ideal filter is

$$\beta_j = \begin{cases} \beta_j^* + \theta & |j| < J \\ 0 & \text{otherwise} \end{cases},$$

where θ is chosen to achieve the condition $B(1) = 0$. See figure 1 (from Baxter, M. and King, R., (1999), .Measuring Business Cycles: Approximate Band-Pass Filters for Economic Time Series., Review of Economics and Statistics, 81, 575-593) for an illustration of how well this works. With $J = 16$, the truncated filter is quite close to the ideal one. In fact, what Baxter-King filter does - it solves

$$\begin{cases} \min_{\{b_{-k}, \dots, b_k\}} \frac{1}{2\pi} \int_{-\pi}^{\pi} \|b(e^{i\omega}) - b^*(e^{i\omega})\|^2 d\omega \\ \text{s.t } b^*(1) = 0 \end{cases}$$

A problem from problem set 1 may give an idea why it can be good.

In addition to the band-pass filter just described, Baxter-King suggest a high-pass filter that removes only a trend and keeps high frequency, and a low-pass filter that leaves only low frequencies.

Remark 10. This is far from the only approach to isolating the business cycle. The simplest approach would be to just detrend GDP. The Hodrick-Prescott filter is another approach.

Remark 11. We developed our theory for stationary processes, but now we're talking about removing trends from non-stationary data. Once we apply these filters, we end up with a stationary series.

Definition 12. *Hodrick-Prescott Filter:* Assume GDP is $x_t = c_t + \tau_t$, a cycle c_t plus a trend τ_t . We choose c_t and τ_t by solving

$$\min_{\tau_t} \left\{ \sum (x_t - \tau_t)^2 + \lambda(\tau_{t+1} - 2\tau_t + \tau_{t-1})^2 \right\}$$

where λ controls how much we penalize the trend for non-smoothness ($\tau_{t+1} - 2\tau_t + \tau_{t-1}$ is like a 2nd derivative). For quarterly data, HP suggests $\lambda = 1600$.

Words of caution

- *Filters are two-sided:* they average over today's, yesterday's and tomorrow's data. If we're interested in whether one series leads (or perhaps causes) another we may mess up the relationship by filtering.
- *Filters should be applied to both data and models:* if we're comparing simulations from a model to the data, we should apply the same filter to both.
- *HP can generate spurious cycles:* Copley and Nason (1995) generated random-walk data, applied HP, and found cycles. BK might also generate spurious cycles. In general, we know the theory of these filters, but we don't know their stochastic properties so well.

MIT OpenCourseWare
<http://ocw.mit.edu>

14.384 Time Series Analysis
Fall 2013

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.