The frequency domain

Tom Holden

http://www.tholden.org/

PhD Macroeconomics, Semester 2

Outline of today's talk

- The frequency domain.

- Filters.

Readings on continuous time processes etc.

- Canova: "Methods for applied macroeconomic research".
  - Section 1.6 covers the frequency domain.
  - Section 3.2 covers filters.

- Christiano-Fitzgerald (2003)
  - Introduces a common filter.
  - Working paper version here: http://www.clevelandfed.org/Research/workpaper/1999/Wp9906.pdf

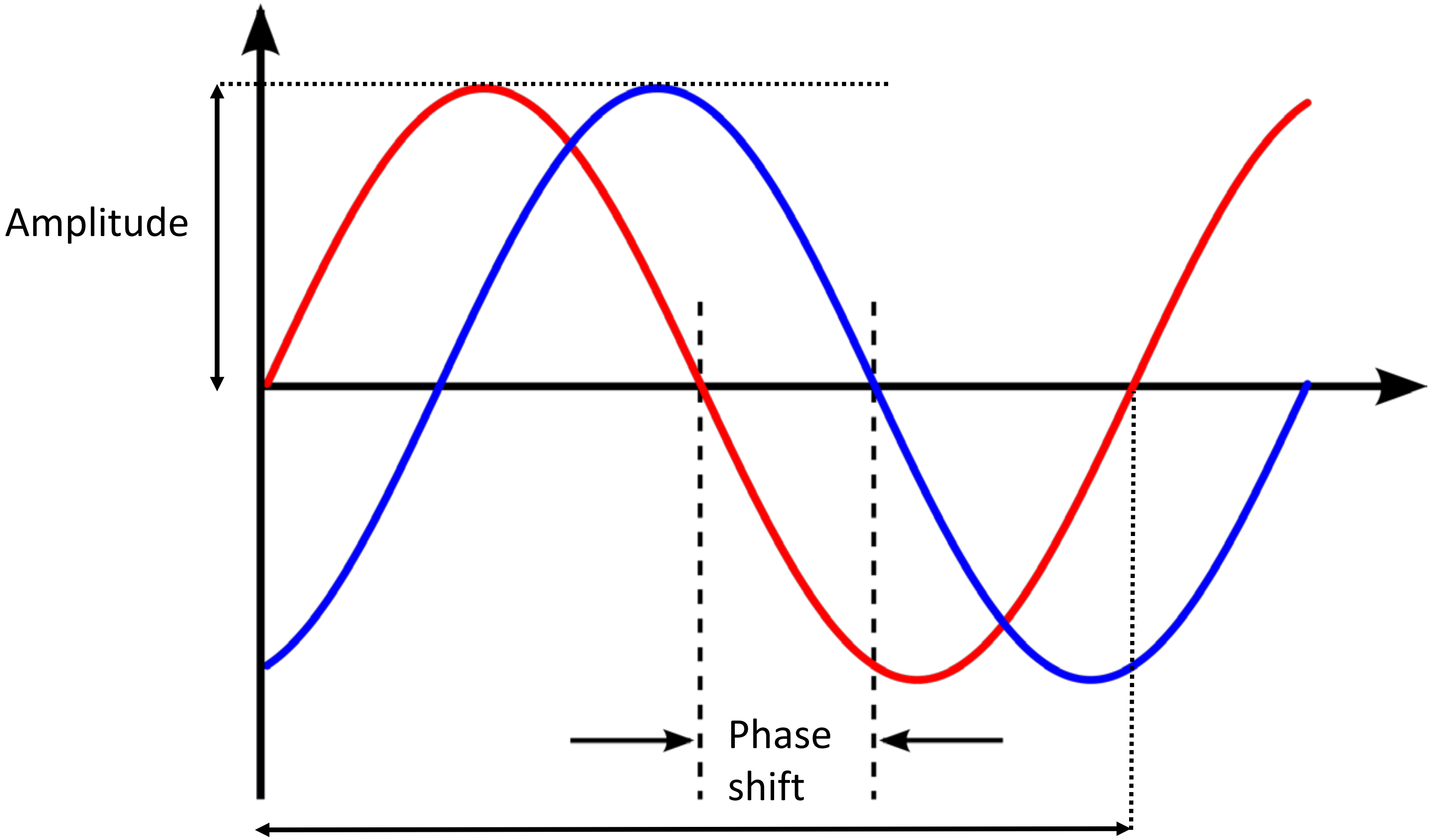
- Wikipedia as needed…

The frequency domain

- We are interested in business “cycles”.

- This suggests that we ought to be concerned with the characteristics of the data in the frequency domain.
  - I.e. we want to know at what frequencies (equivalently: period lengths) is the variance of the data?
  - Low frequency variation (normally defined as cycles of over 50 years) captures very persistent components of the data, such as structural change.
  - Medium frequency variation (normally defined as cycles of 8-50 years) captures growth dynamics.
  - High frequency variation (normally defined as cycles of below 2 years) is driven by seasonal patterns, and noise.
  - Business cycles are what’s left (so normally cycles of 2-8 years).

Frequency and phase



Source: https://en.wikipedia.org/wiki/File:Phase_shift.svg

The Fourier transform on an interval: Introduction (1/2)

- Suppose you have a vector $u \in \mathbb{R}^n$, and you wish to know the length of that vector in a particular direction $v \in \mathbb{R}^n$ (with $\|v\| = 1$), what do you do?
  - You take the inner ("dot") product of u and v, i.e. $\langle u, v \rangle = v'u = \sum_{i=1}^{n} u_i v_i$.
- Recall also that any element of $\mathbb{R}^n$ may be expressed as a linear combination of n basis vectors.
- How do we find the coefficients? We just take the inner product with each basis vector in turn.
- These ideas extend to other vector spaces.

The Fourier transform on an interval: Introduction (2/2)

- Consider the space of all (possibly complex) square integrable functions on the interval $[0,1]$.
- The natural inner product here is $\langle f, g \rangle = \int_0^1 f(x)\overline{g(x)}\,dx$, where $\overline{}$ denotes the complex conjugate.
- The remarkable thing is that this space also has a countably infinite basis, despite the interval $[0,1]$ being uncountable.

- This basis is made up of the functions $x \mapsto e^{2\pi i n x}$ for all $n \in \mathbb{Z}$, where $i = \sqrt{-1}$.
  - Recall that $e^{i\phi} = \cos\phi + i\sin\phi$, so this basis is expressing the function as sums of sines and cosines at different integer frequencies.

- The Fourier transform recovers the coefficients on these basis functions.
  - As you would expect, it takes the form of an inner product of the function of interest with the basis functions.

The Fourier transform on an interval: Details

- Define $e_n: [0, 1] \to \mathbb{C}$ by $e_n(x) = e^{2\pi i n x}$ for all $x \in [0,1]$.
- Carleson's theorem states that for any square integrable function $f: [0,1] \to \mathbb{C}$, and almost all $x \in [0,1]$:

$$f(x) = \sum_{n=-\infty}^{\infty} \langle f, e_n \rangle e_n(x) \,.$$

- I.e. if we define:

$$a_n \coloneqq \langle f, e_n \rangle = \int_0^1 f(x)\overline{e_n(x)}\, dx = \int_0^1 f(x) e^{-2\pi i n x}\, dx \,,$$

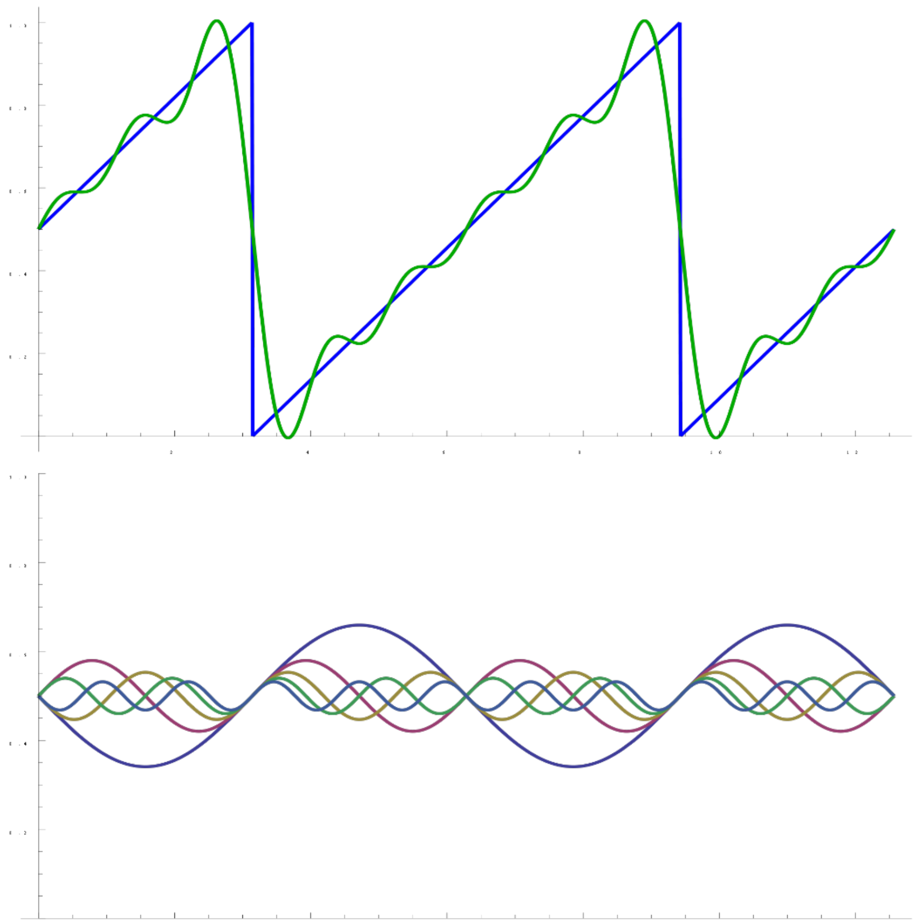
- Then for almost all $x \in [0,1]$:

$$f(x) = \sum_{n=-\infty}^{\infty} a_n e^{2\pi i n x}$$

- The Fourier transform $\mathcal{F}: ([0,1] \to \mathbb{C}) \to (\mathbb{Z} \to \mathbb{C})$ is then given by $\mathcal{F}f = a = (a_n)_{n \in \mathbb{Z}}$, where a_n is given as above.

- The Fourier transform is invertible, with $\mathcal{F}^{-1}a = \sum_{n=-\infty}^{\infty} a_n e_n$.

The Fourier transform on an interval: Example



Source: https://en.wikipedia.org/wiki/File:Sawtooth_Fourier_Analysys.svg

The Fourier transform in discrete time

- To complete our suite of definitions, we need to define the Fourier transform for discrete time processes.
- Recall that the Fourier transform of a function on the unit interval was a sequence. It shouldn't be surprising then that the Fourier transform of a sequence is a function on the unit interval.

- In this case, we define $\mathcal{F}: (\mathbb{Z} \to \mathbb{C}) \to ([0,1] \to \mathbb{C})$ by:

$$\mathcal{F}(a)(\xi) = \sum_{n=-\infty}^{\infty} a_n e^{-2\pi i n \xi} .$$

  - The only difference to the inverse transform given previously is the negative sign, really just a matter of convention!

- As before, $\mathcal{F}$ is invertible, in the sense that $\left(\mathcal{F}^{-1}\big(\mathcal{F}(a)\big)\right)_n = a_n$ for all n, where:

$$\mathcal{F}^{-1}(g)(x) = \int_0^1 g(\xi) e^{2\pi i x \xi} \, d\xi .$$

The Fourier transform in continuous time

- The Fourier transform may also be defined for functions on the real line (e.g. in continuous time).
- In this case, we define $\mathcal{F}: (\mathbb{R} \to \mathbb{C}) \to (\mathbb{R} \to \mathbb{C})$ by:

$$\mathcal{F}(f)(\xi) = \int_{-\infty}^{\infty} f(x) e^{-2\pi i \xi x} \, dx \, .$$

- If f and $\mathcal{F}(f)$ are absolutely integrable, then $\mathcal{F}$ is invertible, in the sense that $\mathcal{F}^{-1}\big(\mathcal{F}(f)\big)(x) = f(x)$ for almost all x, where:

$$\mathcal{F}^{-1}(g)(x) = \int_{-\infty}^{\infty} g(\xi) e^{2\pi i x \xi} \, d\xi \, .$$

- Often the Fourier transforms of processes are much simpler than the original one, so it can be much easier to prove results if the Fourier transform of both sides is taken first.

The Dirac Delta "function" and its Fourier transform

- Define the "function" (actually a measure) $\delta(x)$ by the property:
  - For any function $f(x)$, $\int_{-\infty}^{\infty} f(x)\delta(x)\,dx = f(0)$.
  - You can think of δ as the p.d.f. of a normal distribution with mean 0 and infinitesimal standard deviation. Thus $\delta(0) = \infty$, and $\delta(x) = 0$ for $x \neq 0$.

- This is useful for looking at various degenerate cases, as often, δ can be treated as if it were really a function.

- The Fourier transform can be extended to such measures.

- In particular, it turns out that:
  - $\mathcal{F}(1)(\xi) = \delta(\xi)$.
  - $\mathcal{F}(\delta)(\xi) = 1$.

The spectral density

- Provides an answer to the question: "at what frequencies is the variance of the data?"

- Two equivalent definitions, for a weakly stationary process X_t (either in continuous or discrete time!):
  1. $S_{XX}(\omega) = \mathbb{E}\left|\mathcal{F}(X - \mathbb{E}X)\left(\frac{\omega}{2\pi}\right)\right|^2$. (The expectation of the squared modulus of the Fourier transform of the demeaned process.)
  2. $S_{XX}(\omega) = \mathcal{F}(\gamma_X)\left(\frac{\omega}{2\pi}\right)$, where γ_X is the ACF of X_t. (The Fourier transform of the ACF.)

- Since it is based on the squared Fourier transform of the process, all information about the phase of the signal is lost in the spectral density.
- $\int_{\underline{\omega}}^{\overline{\omega}} S_{XX}(\omega)\, d\omega$ is proportional to variance at frequencies between $\underline{\omega}$ and $\overline{\omega}$.

White noise

- Three definitions:

1. White noise is a hypothetical process with an auto-covariance function given by the Dirac Delta function.
2. White noise is a hypothetical process with constant spectral density 1.
   - By analogue to white light, which is a mix of all frequencies.
3. White noise is the hypothetical derivative of the undifferentiable Wiener process, W_t.

- The properties of the Fourier transform of the Dirac Delta function imply that the first two definitions are equivalent.

The spectral density of an Ornstein-Uhlenbeck process: Direct approach

- Take the Ornstein-Uhlenbeck process: $Z_t = \mu + \sigma \int_0^\infty e^{-\theta s}\, dW_{t-s}$.
- Let γ be its ACF. I.e., assuming $\tau > 0$:

$$\begin{aligned}
\gamma_Z(\tau) &= \mathbb{E}[(Z_t - \mathbb{E}Z_t)(Z_{t-\tau} - \mathbb{E}Z_{t-\tau})] \\
&= \sigma^2 \mathbb{E}\left[\left(\int_{s=0}^\infty e^{-\theta s}\, dW_{t-s}\right)\left(\int_{s=0}^\infty e^{-\theta s}\, dW_{t-\tau-s}\right)\right] \\
&= \sigma^2 \mathbb{E}\left[\left(\int_{-\infty}^{t} e^{-\theta(t-u)}\, dW_u\right)\left(\int_{-\infty}^{t-\tau} e^{-\theta(t-\tau-u)}\, dW_u\right)\right] \\
&= \sigma^2 \int_{-\infty}^{t-\tau} e^{\theta\tau - 2\theta(t-u)}\, du \text{ [By the Ito Isometry]} = \frac{\sigma^2 e^{-\theta\tau}}{2\theta}.
\end{aligned}$$

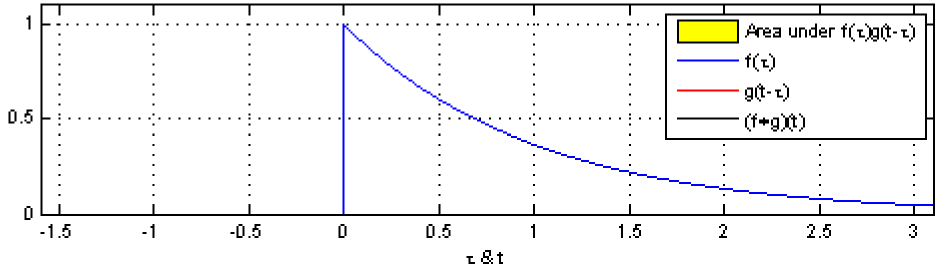
- Then:

$$\begin{aligned}
S_{ZZ}(\omega) = \mathcal{F}(\gamma_Z)\left(\frac{\omega}{2\pi}\right) &= \int_{-\infty}^{0} \frac{\sigma^2 e^{\theta\tau}}{2\theta} e^{-i\omega\tau}\, d\tau + \int_0^\infty \frac{\sigma^2 e^{-\theta\tau}}{2\theta} e^{-i\omega\tau}\, d\tau \\
&= \frac{\sigma^2}{2\theta}\left[\frac{1}{\theta - i\omega} + \frac{1}{\theta + i\omega}\right] \\
&= \frac{\sigma^2}{2\theta}\left[\frac{\theta + i\omega + \theta - i\omega}{(\theta - i\omega)(\theta + i\omega)}\right] = \frac{\sigma^2}{\theta^2 + \omega^2}
\end{aligned}$$

The convolution theorem

- This theorem captures one of the nicest properties of the Fourier transform.
- The convolution of two functions f and g is denoted $f * g$ and is defined by:

$$(f * g)(x) = \int_{-\infty}^{\infty} f(y)g(x-y)\,dy$$



Source: https://en.wikipedia.org/wiki/File:Convolution_of_spiky_function_with_box2.gif

- The convolution theorem states that for almost all ξ:

$$\mathcal{F}(f * g)(\xi) = \mathcal{F}(f)(\xi)\mathcal{F}(g)(\xi).$$

- Simple proof: http://mathworld.wolfram.com/ConvolutionTheorem.html
- The convolution theorem also holds in discrete time (with convolution defined by a sum rather than an integral).

The spectral density of an Ornstein-Uhlenbeck process: Lazy approach

- We may define an operator D which takes the time derivative of a differentiable process, so $Df(t) = \frac{df(t)}{dt}$.
- If we are careful, we may extend this operator to continuous time stochastic processes, even though they may not be differentiable.
  - $g(t, X_t)DX_t$ on its own will not make sense, but we can define $\int_a^b g(t, X_t)DX_t\, dt = \int_a^b g(t, X_t)\, dX_t$.

- Given this definition, if $Z_t = \mu + \sigma \int_{s=0}^{\infty} e^{-\theta s}\, dW_{t-s}$, and $h(s) = \mathbb{1}(s > 0)e^{-\theta s}$ then $Z_t = \mu + \sigma(h * DW)(t)$.
- Hence, by the convolution theorem and the properties of white noise:

$$\mathbb{E}\left|\mathcal{F}(Z_{\cdot} - \mathbb{E}Z_{\cdot})\left(\frac{\omega}{2\pi}\right)\right|^2 = \sigma^2 \left|\mathcal{F}(h)\left(\frac{\omega}{2\pi}\right)\right|^2 \mathbb{E}\left|\mathcal{F}(DW_{\cdot})\left(\frac{\omega}{2\pi}\right)\right|^2 = \sigma^2 \left|\mathcal{F}(h)\left(\frac{\omega}{2\pi}\right)\right|^2.$$

$$= \sigma^2 \left|\int_{-\infty}^{\infty} \mathbb{1}(s > 0)e^{-\theta s}e^{-i\omega s}\, ds\right|^2 = \sigma^2 \left|\frac{1}{\theta + i\omega}\right|^2 = \frac{\sigma^2}{\theta^2 + \omega^2}.$$

The spectral density of an Ornstein-Uhlenbeck process: Interpretation

- The convolution theorem allowed us to write the spectral density as the product of the spectral density of white noise, and the squared Fourier transform of some deterministic function.

- In effect then, we are filtering out the frequencies we don't like from the original white noise.

- When we look at frequency domain filters later, this is exactly how they will be defined. To filter the data, we will transform it into the frequency domain, and then multiply it pointwise by some function.

- As a result of their filtering behaviour, processes with an $\mathrm{MA}(\infty)$ representation are often termed linear filters.

The spectral density of an arbitrary linear filter in discrete time

- Suppose x_t is a weakly stationary process, c is the polynomial $c(\lambda) = \sum_{s=0}^{\infty} c_s \lambda^s$, where $\sum_{s=0}^{\infty} c_s^2 < \infty$ and $y_t = \mu + c(L)x_t$.

- Then, if we define $h(s) = \mathbb{1}(s \geq 0)c_s$:

$$
\begin{aligned}
S_{yy}(\omega) &= \mathbb{E}\left|\mathcal{F}\left(t \mapsto \sum_{s=0}^{\infty} c_s x_{t-s}\right)\left(\frac{\omega}{2\pi}\right)\right|^2 \\
&= \mathbb{E}\left|\mathcal{F}(h * x.)\left(\frac{\omega}{2\pi}\right)\right|^2 \\
&= \left|\mathcal{F}(h)\left(\frac{\omega}{2\pi}\right)\right|^2 \mathbb{E}\left|\mathcal{F}(x.)\left(\frac{\omega}{2\pi}\right)\right|^2 \\
&= \left|\sum_{s=0}^{\infty} c_s e^{-is\omega}\right|^2 S_{xx}(\omega) \\
&= \left|c\left(e^{-i\omega}\right)\right|^2 S_{xx}(\omega) = c\left(e^{-i\omega}\right)c\left(e^{i\omega}\right)S_{xx}(\omega)
\end{aligned}
$$

The spectral density of an $\mathrm{ARMA}(p,q)$ process.

- Suppose $\Phi_p(L)y_t = \mu + \Theta_q(L)\sigma\varepsilon_t$, where Φ_p and Θ_q are polynomials of degree p and q respectively, and $\varepsilon_t \sim \mathrm{NIID}(0,1)$.
- Then: $y_t = \frac{\mu}{\Phi_p(L)} + \frac{\Theta_q(L)}{\Phi_p(L)}\sigma\varepsilon_t$.
- Applying the previous result we have:

$$S_{yy}(\omega) = \sigma^2 \frac{\Theta_q\left(e^{-i\omega}\right)\Theta_q\left(e^{i\omega}\right)}{\Phi_p(e^{-i\omega})\Phi_p(e^{i\omega})} S_{\varepsilon\varepsilon}(\omega).$$

- Just as in the continuous time case, $S_{\varepsilon\varepsilon}(\omega) = 1$.
- Hence:

$$S_{yy}(\omega) = \sigma^2 \frac{\Theta_q\left(e^{-i\omega}\right)\Theta_q\left(e^{i\omega}\right)}{\Phi_p(e^{-i\omega})\Phi_p(e^{i\omega})}.$$

- For example, if $p = q = 1$, and $\Phi_1(\lambda) = 1 - \phi\lambda$ and $\Theta_1(\lambda) = 1 + \theta\lambda$:

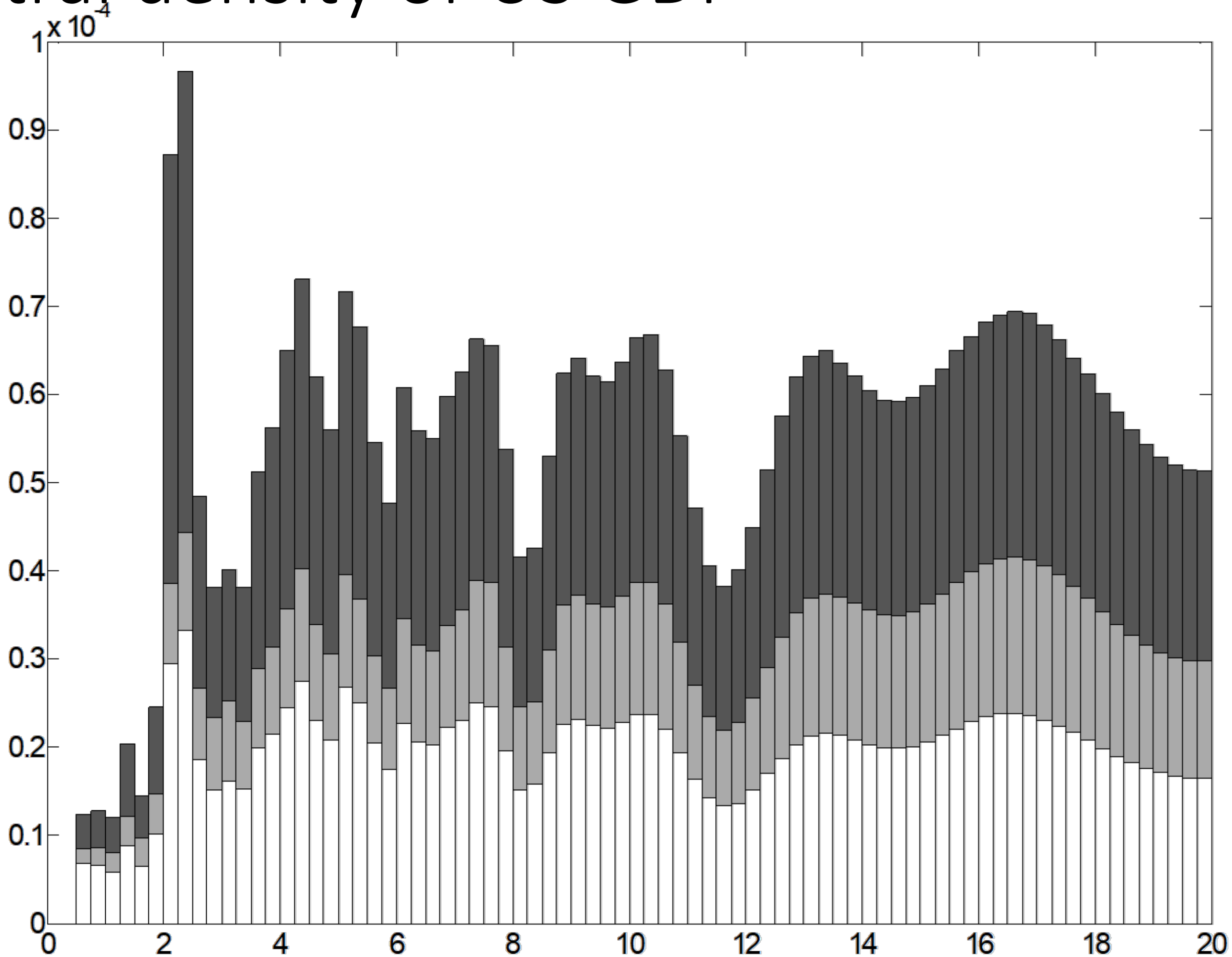
$$S_{yy}(\omega) = \sigma^2 \frac{\left(1 + \theta e^{-i\omega}\right)\left(1 + \theta e^{i\omega}\right)}{(1 - \phi e^{-i\omega})(1 - \phi e^{i\omega})} = \sigma^2 \frac{1 + 2\theta\cos\omega + \theta^2}{1 - 2\phi\cos\omega + \phi^2}.$$

Estimating spectral densities

- One (parametric) method is to fit an $\mathrm{ARMA}(p, q)$ then use the previous formula to get an estimate of the spectrum.

- Most non-parametric estimates are based around the sample auto-covariance function.
  - The "periodogram" is the Fourier transform of the sample auto-covariance function. This is asymptotically unbiased, but unfortunately it is inconsistent, intuitively because you would need an infinite amount of data to get the variance at frequency 0.
  - As is standard in non-parametric econometrics, to derive a consistent estimator you must smooth the data via some kernel. In spectral density estimation, this smoothing may be applied either to the ACF, or to its Fourier transform.
  - Quite difficult in practice, and getting reasonable standard errors is even harder. (I spent a long time a few years ago trying to get a reasonable spectral density estimate for US real GDP per capita.)

Spectral density of US GDP



The horizontal axis gives the period length in years here, not the frequency.
(This is non-standard, but makes things clearer.)

Business cycle filters

- As may be seen by the previous plot, macro time series have a lot of variance at frequencies well below business cycle frequencies.

- Thus, if we are going to assess the performance of a model designed to match just the business cycle, we might like to filter out other frequencies prior to comparing the model to the data.

- In the early literature, this was done using the Hodrick-Prescott (1997) filter, which is in the time domain.

- The modern literature uses the Christiano-Fitzgerald (2003) filter, or some other frequency domain one instead.

The Hodrick-Prescott (HP) filter

- Suppose x_t is some time series of length T.
- The HP filtered version of x_t is the sequence $x_t - \tau_t$, where τ_t is the "HP-trend", which is the solution to the following problem:

$$\min_{\tau_1,\ldots,\tau_T} \left[\sum_{t=1}^{T} (x_t - \tau_t)^2 + \lambda \sum_{t=2}^{T-1} (\Delta\tau_{t+1} - \Delta\tau_t)^2 \right].$$

  - where λ is some constant, usually, $\lambda = 1600$ for quarterly data.

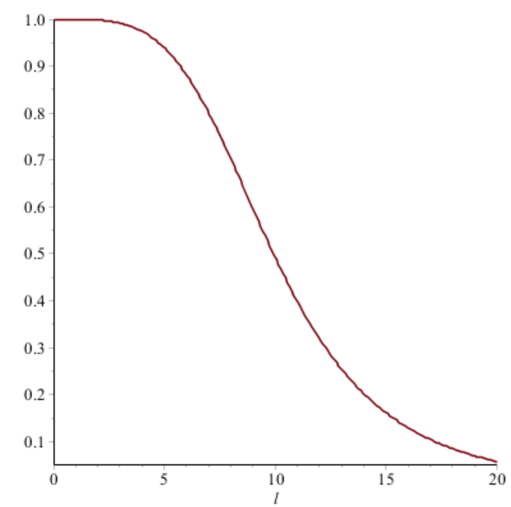
- Problems:
  - Since this is a time domain filter, there's no guarantee it's going to recover the frequencies we're interested in.
  - The filter is non-causal, i.e. the filtered observation at t depends on the source data at $t + 1$, $t + 2$, etc.
  - The filter suffers from "end-point bias", with the first and last observations having large impacts on the estimated trend.

Frequency domain filters

- An ideal filter would attenuate frequencies by some desired amount.
  - For example, if we're interested in the business cycle, we might like a "band-pass filter" which completely cut out all frequencies with periods below two years or above eight years, while leaving frequencies in between unaffected.
- More generally, a filter is defined by its frequency response function.
  - This gives the attenuation at a specified frequency, where a value of 0 means full attenuation, and a value of 1 means none.
- King and Rebelo (1993) showed that the frequency response of the HP-filter at frequency ω is given by:

$$\frac{4\lambda(1-\cos\omega)^2}{1+4\lambda(1-\cos\omega)^2}$$

- This is plotted on the right, with period length in years on the horizontal axis.



The Christiano-Fitzgerald (CF) band-pass filter

- One algorithm for band-pass filtering the data is to take its Fourier transform, then take the pointwise product of this with the desired frequency response function.
  - If you have infinite data, this works perfectly.
  - Unfortunately, in finite samples it performs poorly, as the temporal truncation is like multiplying the series by a box function. Since the box function has a Fourier transform of the form $\frac{\sin \omega}{\omega}$, applying the Fourier transform to a finite sample is like convolving the data in the frequency domain with $\frac{\sin \omega}{\omega}$.
- Thus, in order to produce a well performing filter, we need a way of extending a finite sample forwards and backwards in time.
- The idea of the CF filter is to approximate the series by a random walk outside of the observed window.
  - For many macro time series, this will be a good approximation.
  - See the paper for details.

Behaviour of the Christiano-Fitzgerald filter

- The standard version of the CF filter is asymmetric (i.e. its frequency response is not an even function), so it may introduce phase shifts.
  - Phase shifts are highly undesirable in macro contexts, as they will disrupt inference about which variables lead which other variables.
- However, CF also provide a symmetric version.
  - Matlab code for all versions is here: http://www.clevelandfed.org/research/models/bandpass/bpassm.txt
- Lee and Steehouwer (2012) show that the CF filter also tends to perform poorly towards the ends of the interval, as shown by the figure below from their paper:

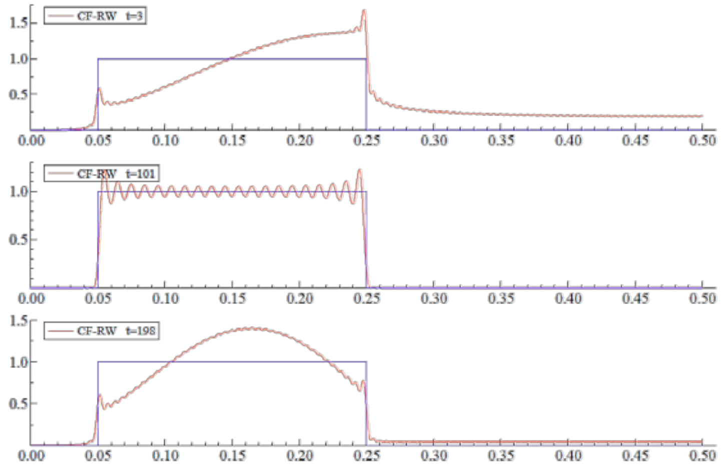


Figure 2.2: Squared gain of ideal and CF-RW filter with $T = 201$ observations at various times t, with pass band [1/20, 1/4] (frequency as number of cycles per period).

Conclusion and recap

- Linear processes can be thought of as filters on white noise.
- The convolution theorem is great!
- Care must be taken when filtering the data.