SVARs & identification, and advanced time series (continuous time processes and the frequency domain)

Tom Holden

http://www.tholden.org/

PhD Macroeconomics, Semester 2

Outline of today's talk

- SVARs.
- Identification methods.
- Continuous time stochastic processes.
- The frequency domain.
- Filters.

Reading on SVARs and identification

- Canova: “Methods for applied macroeconomic research”.
  - Section 4.5 covers Identification and SVARs.
  - Section 10.3 covers this in a Bayesian context.

- Wikipedia, as needed for basic results in linear algebra.
  - Reading all of the pages in this category would be a good start: https://en.wikipedia.org/wiki/Category:Matrix_decompositions

- Christiano, Eichenbaum and Evans (2003):
  - A classic paper which you ought to be familiar with.
  - http://benoitmojon.com/pdf/Christiano%20%20Eichenbaum%20Evans%202005%20JPE.pdf

Readings on continuous time processes etc.

- Canova: “Methods for applied macroeconomic research”.
  - Chapter 1.6 and 3 covers the frequency domain and filters.
- Cochrane (2012):
  - Nice review of continuous time stochastic processes, with a macro slant.
  - http://faculty.chicagobooth.edu/john.cochrane/research/papers/continuous_time_linear_models.pdf
- Christiano-Fitzgerald (2003)
  - Introduces a common filter.
  - Working paper version here: http://www.clevelandfed.org/Research/workpaper/1999/Wp9906.pdf
- Wikipedia as needed…

Structural VARs: Motivation (1/2)

- We would like to know what the effects of (say) an unexpected increase in monetary policy is.
- But a change in monetary policy will produce changes in other variables within the same time period.
- Conversely, exogenous shocks to other variables will produce automatic reactions from monetary policy.
  - E.g. a Taylor Rule.
- Thus, if we see that interest rates were (say) tighter than was expected yesterday, we do not know if this was due to a change in policy or if it was an endogenous reaction to other changes in the economy.
  - A standard VAR tells us nothing about the effects of changes in policy!

Structural VARs: Motivation (2/2)

- Furthermore, even after contemporaneous responses of one variable to another have been taken into account, there may still be correlations in the shocks.
  - For example, an exogenous increase in rainfall may both decrease labour supply holding fixed the wage, and increase labour demand holding fixed the wage. Thus in a VAR in which rainfall is omitted, it would show up as both a labour supply and a labour demand shock.
  - No way of knowing how much of this variance component due to rainfall should be assigned to supply, and how much should be assigned to demand.
  - However, this really reflects a failure of the model (omitting an observable variable).

  - Alternatively, some variables may respond directly to structural shocks to other variables.
  - In macroeconomic terms this is rather implausible, as shocks are generally not observed directly, and if they are observed, they're generally only observed by the agent that experiences the shock.
  - Nonetheless, in a few rare cases this may be justified.

Structural VARs: Definition

- This suggests the following structural representation:

$$x_t = c + a_0 x_t + a_1 x_{t-1} + \cdots + a_p x_{t-p} + u_t + b_0 u_t$$

  - where both a_0 and $b_0$have a zero diagonal and where $u_t \sim \mathrm{WNIID}(0, \Sigma_u)$, with Σ_u diagonal.

- Then:

$$(I - a_0) x_t = c + a_1 x_{t-1} + \cdots + a_p x_{t-p} + (I + b_0) u_t$$

- Then if we define $A \coloneqq I - a_0$ and $B \coloneqq I + b_0$:

$$x_t = A^{-1} c + A^{-1} a_1 x_{t-1} + \cdots + A^{-1} a_p x_{t-p} + A^{-1} B u_t.$$

- Compare this to our previous reduced form:

$$x_t = \mu + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + \varepsilon_t, \qquad \varepsilon_t \sim \mathrm{WNIID}(0, \Sigma_\varepsilon)$$

- Matching terms gives:

$$A\mu = c, \qquad A\phi_1 = a_1, \qquad \ldots, \qquad A\phi_p = a_p, \qquad A\varepsilon_t = B u_t,$$
$$A \Sigma_\varepsilon A' = B \Sigma_u B'$$

Structural VARs: Basic identification (1/2)

- Our hope is to be able to use some prior restrictions on A and B (or equivalently a_0 and b_0), in order to solve for Σ_u in the equation $A\Sigma_\varepsilon A' = B\Sigma_u B'$.
- We know A and B have a unit diagonal, and that Σ_u is zero everywhere except the diagonal.
- If we knew $A\Sigma_\varepsilon A'$ could we at least work out B and Σ_u?
  - No, not uniquely, without additional information.
  - By the Cholesky decomposition, there exists a lower triangular matrix L such that $A\Sigma_\varepsilon A' = LL'$. So one candidate solution is $B \coloneqq L(\operatorname{diag}\operatorname{diag} L)^{-1}$, and $\Sigma_u \coloneqq (\operatorname{diag}\operatorname{diag} L)^2$.
  - But let U be any real orthogonal matrix. Then $A\Sigma_\varepsilon A' = (LU)(LU)'$ too. Thus $B \coloneqq LU(\operatorname{diag}\operatorname{diag} LU)^{-1}$, and $\Sigma_u \coloneqq (\operatorname{diag}\operatorname{diag} LU)^2$ is another solution.
  - The space of all $n \times n$ orthogonal matrices is $\frac{n(n-1)}{2}$ dimensional, so this is the number of restriction we need on B if we already know $A\Sigma_\varepsilon A'$.
  - This may be seen directly from noting that in the equation $RR' = S$, with S symmetric, the equations above the diagonal are identical to those below.

Structural VARs: Basic identification (2/2)

- In practice B is almost always assumed to be equal to the identity matrix, for the reasons I gave previously.
  - If it's not, it reflects either strange informational assumptions, or omitted variables.
- So with B known, can we pin down A without additional assumptions?
  - No. Much as before, the equation $A\Sigma_\varepsilon A' = B\Sigma_u B'$ has $\frac{n(n-1)}{2}$ free parameters with B known.
  - So this is the number of assumptions we need to make on A.
- A common assumption is that A is lower triangular, which gives the required $\frac{n(n-1)}{2}$ restrictions.
  - This means a_0 is strictly lower triangular, implying a "causal ordering" on the variables.
  - The variable ordered first is assumed to have no contemporaneous response to later variables.
  - The one ordered second just responds to the first contemporaneously, but no others. Etc. etc. till…
  - The one ordered last responds contemporaneously to all variables.
- Exercise: how many restrictions on A and B do we need when neither are known?

Reporting SVAR results: Impulse responses (1/2)

- What do we do once we've decided upon an identification?
- Normally, we are primarily interested in the response of the economy to some shock.
  - So, suppose there's a shock of one standard deviation to variable i in period 0, and then from then on, no further shocks hit the economy.
  - I.e., suppose that $u_{0,i} = 1$ for some $i \in \{1, \dots, n\}$, but that $u_{t,j} = 0$ for all $t \in \mathbb{N}$ and all $j \in \{1, \dots, n\}$, unless $t = 0$ and $j = i$.
  - Given these assumptions on the shocks, we can simulate the SVAR model and plot the responses of the variables of interest.
  - The result of this is a so-called "impulse response".
    - Often these are plotted relative to the variable's mean.

Reporting SVAR results: Impulse responses (2/2)

- In non-linear models, there are various possible definitions of an impulse response.
  - One is $\mathbb{E}\left[x_t \middle| u_{0,i} = 1\right]$.
  - Another is $\mathbb{E}\left[x_t \middle| u_{0,i} = \tilde{u}_{0,i} + 1\right]$, where $\tilde{u}_{0,i}$ has the same distribution as that of $u_{t,i}$.
    - Some authors also condition on the initial state in these expectations. (Dynare does not.)

- Exercise: prove that these definitions for non-linear models both agree with the standard definition in a standard linear SVAR.

Other identification methods: AB restrictions

- Causal orderings are deeply implausible. Most variables have some contemporaneous effect on most other variables.
- Indeed, many variables have strong anticipatory effects on other variables.
  - If a shock to another variable is expected in future (and the econometricians data-set is insufficient to pick this up) then shocks (observed) tomorrow might have an effect on variables today.

- Other restrictions on the A and B matrices based on theory are often as bad, for basically the same reason.

- The Blanchard and Perotti (2002) approach uses micro data to estimate some parameters of the A matrix in a fiscal policy context.
  - Their argument is based upon government taking more than quarter to respond, so is little better than the causal ordering approach.

Other identification methods: Sign restrictions

- The sign restriction approach (Uhlig 2005) effectively places a flat prior over the space of all orthogonal rotation matrices, then truncates this prior to zero in areas where the model generates “the wrong results” in some sense.
- “Wrong” is usually defined in terms of the sign of the impulse response to a certain shock at a certain point in time.
  - May end up assuming what it wants to prove. For example, causal ordering identification of monetary policy shocks often produces “price puzzles”, with increasing interest rates increasing inflation.
  - Assuming away price puzzles begs the question of whether these are real features or not.
  - It’s certainly easy to generate price puzzles in theoretical models, so the theoretical grounds for excluding them are very weak.
- If interpreted classically, sign restrictions only produce set identification, not point identification. (see Moon, Schorfheide and Granziera 2013).
  - Following identification via sign restrictions, there is no such thing as “the” estimated impulse response.
  - Rather, the estimator provides a band of impulse responses, even with infinite data.
  - Finite sample parameter uncertainty produces even larger bands.

Other identification methods: Narrative evidence

- Pioneered by Romer and Romer (1989), who use the text of FOMC meetings to identify times when policy makers intended to use contractionary policy to bring down inflation.
  - Later work has tightened the definition of a monetary shock.
- In fiscal contexts, Ramey and Shapiro (1998) performed a similar analysis using military build-ups.

- One difficulty with this approach is that hand selected shocks will always "smell funny".
  - Some recent researchers have ameliorated this via using automated textual analysis.
- Another problem is that it's not always clear that the narrative analysis procedure really succeeds in purging all endogeneity.

Other identification methods: Long-run restrictions

- While there's a lot of debate about how the economy evolves in the short-run, there's a lot more consensus about the long-run effects of various shocks.
  - E.g. only a technology shock increases GDP per capita in the long-run. Monetary shocks are neutral for all variables in the long-run. Etc.

- Blanchard and Quah (1989) exploit this for identification.
  - It is a bit like a sign restriction at $t = \infty$, but since they are imposing exact coefficients for the long-run response they get point, not set, identification.
  - Furthermore, no simulation is necessary, since in a reduced form VAR with $\mu = 0$, the long-run response to a one off shock is as follows:

$$x_t = \lim_{t\to\infty} \begin{bmatrix} I_{n\times n} & 0_{n\times(p-1)n} \end{bmatrix} \begin{bmatrix} \begin{bmatrix} \phi_1 & \cdots & \phi_{p-1} \end{bmatrix} & \phi_p \\ I_{(p-1)n\times(p-1)n} & 0_{(p-1)n\times n} \end{bmatrix}^t \begin{bmatrix} \varepsilon_0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

  - By diagonalising the large matrix we may find the directions in which the model's response is permanent. (Exercise.)

SVARs and identification in practice

- There is a huge literature looking at the responses of monetary and fiscal shocks.
- Results vary wildly depending on which identification method is used, though there is more consensus about monetary shocks than fiscal ones.
  - For example, in a cross country study Iletzki, Mendoza and Vegh (2013) find basically zero fiscal multipliers in developed, open economies, and Ramey's narrative based work finds at most moderate multipliers, around one.
  - On the other hand Perotti continues to find large multipliers.
- The correct response is broad distrust of most VAR identification methods.
- In any case, it is unclear why we should care about fiscal multipliers.
  - The fact that government expenditure increases GDP more than one for one tells us nothing about whether this is good for welfare.
  - In fact, in most modern macro models that generate large multipliers, expansionary fiscal policy is unambiguously bad for welfare.

Continuous time stochastic processes

- The stochastic processes we looked at in the first lecture were random variables taking their value from the vector space of sequences (i.e. functions $\mathbb{Z} \to \mathbb{R}$).

- In some circumstances, it is easier to work in continuous time, i.e. with random variables taking their value from the vector space of functions $\mathbb{R} \to \mathbb{R}$.
  - This is the standard in finance.
  - It's also increasingly more common in macro, and we'll look at some continuous time DSGE models later in the course.

The Wiener process

- The Wiener process (aka “standard Brownian motion”) is the building blocks of most continuous time stochastic processes.
- You might find it helpful to think of the Wiener process as the continuous time analogue of a random walk.
- The process, W_t is characterised by the following properties:
  1. $W_0 = 0$.
  2. W_t is almost surely everywhere continuous.
  3. If $s_1, s_2, \dots, s_{n+1}, t_1, t_2, \dots, t_n \in \mathbb{R}^+$ satisfy $0 < s_i < t_i \leq s_{i+1}$ for all $i \in \{1, \dots, n\}$, then $W_{t_i} - W_{s_i}$ is independent of $W_{t_j} - W_{s_j}$ for all $i \neq j$.
  4. If $0 \leq s < t$, then $W_t - W_s \sim \mathrm{N}(0, t - s)$.

- The process $\mu t + \sigma W_t$ is called a Wiener process with drift μ and infinitesimal variance σ^2.
  - Exercise: what is the unconditional distribution of this process at t?

The Itō integral

- We would often like to work with processes with time varying drift and time varying infinitesimal variance.
- Scaling the Wiener process by a time varying amount will not work, as this will change both the level of the process and its future infinitesimal variance.
- In some loose sense then, we need to "differentiate" the process, scale it, and then integrate back.
- However, the Wiener process is not differentiable.

- Itō defined a new integral (with different integration laws) in order to tackle this.
  - It allows us to integrate a function times a kind of "derivative" of the Wiener process.
- In particular, if:
  - for all $n \in \mathbb{N}$, π_n is an increasing sequence of length $n+1$, with $\pi_{n,0} = 0$ and $\pi_{n,n} = t$,
  - $\lim_{n\rightarrow\infty} \max_{i\in\{1,\ldots,n\}} \left|\pi_{n,i} - \pi_{n,i-1}\right| = 0$,
  - W_t is a Wiener process, and X_t is another continuous time stochastic process that is left-continuous and locally bounded,
  - then we define:

$$\int_0^t X_t \, dW_t \coloneqq \operatorname*{plim}_{n\rightarrow\infty} \sum_{i\in\{1,\ldots,n\}} X_{\pi_{n,i-1}} \left(W_{\pi_{n,i}} - W_{\pi_{n,i-1}}\right).$$

Drift diffusion processes

- Processes used in finance (and continuous time macro) often take the form:
$$X_t = X_0 + \int_0^t \mu(X_u, u)\,du + \int_0^t \sigma(X_u, u)\,dW_u\,.$$
  - The first integral here is a standard integral, the second is an Itō one!
  - The function $\mu(X_t, t)$ controls the drift of the process at t.
  - The function $\sigma(X_t, t)$ controls the infinitesimal variance at t.

- In practice, this expression is usually written in the more compact "stochastic differential equation" form:
$$dX_t = \mu(X_t, t)\,dt + \sigma(X_t, t)\,dW_t.$$
- However, it is important to remember that the latter expression is just a shorthand for the former.

Itō's lemma

- Itō's lemma is an equivalent of the chain rule for continuous time stochastic processes.
- Suppose:

$$dX_t = \mu_t \, dt + \sigma_t \, dW_t.$$

- Then for any twice differentiable function $f: \mathbb{R}^2 \to \mathbb{R}$, Itō's lemma states:

$$df(t, X_t) = \left(f_1 + \mu_t f_2 + \frac{1}{2} \sigma_t^2 f_{22} \right) dt + \sigma_t f_2 \, dW_t$$

- For example, let $Y_t = \exp X_t$, then:

$$dY_t = d \exp X_t = \left(\mu_2 Y_t + \frac{1}{2} \sigma_t^2 Y_t \right) dt + \sigma_t Y_t \, dW_t.$$

Exercise: Apply Itō's lemma to $Z_t \coloneqq a(t) + b(t) X_t$.

Ornstein-Uhlenbeck processes

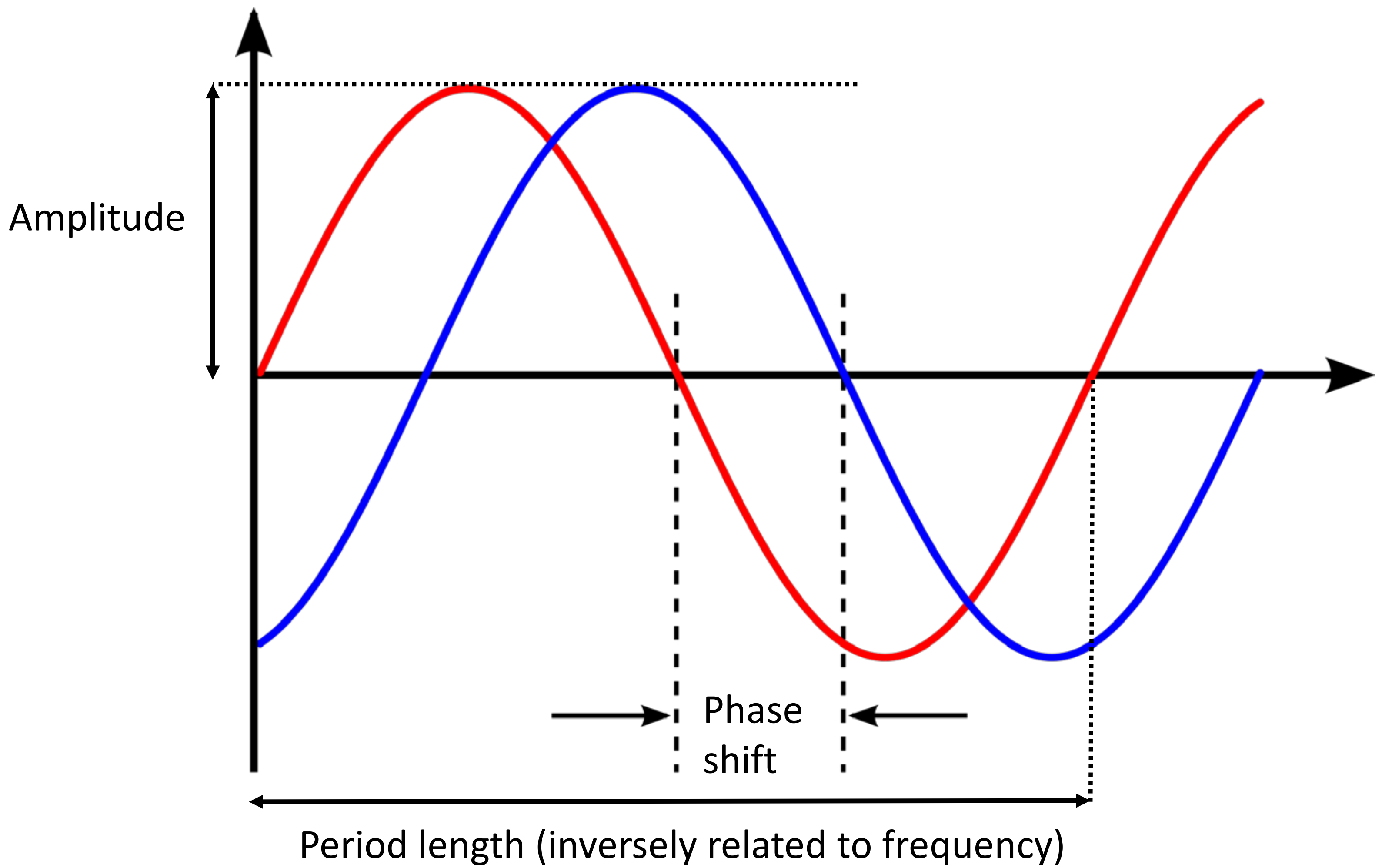
- Ornstein-Uhlenbeck processes are the continuous time equivalent of $\mathrm{AR}(1)$ processes.
  - Recall for later that the $\mathrm{AR}(1)$ process $x_t = (1-\rho)\mu + \rho x_{t-1} + \sigma\varepsilon_t$ has an $\mathrm{MA}(\infty)$ representation $x_t = \mu + \sigma \sum_{s=0}^{\infty} \rho^s \varepsilon_{t-s}$.

- Ornstein-Uhlenbeck processes are solutions to the s.d.e.:
$$dX_t = \theta(\mu - X_t)\,dt + \sigma\,dW_t.$$
- To find their properties, first define $Y_t = e^{\theta t}X_t$, then, by Itō's lemma:
$$\begin{aligned} dY_t = d\left(e^{\theta t}X_t\right) &= \left(\theta Y_t + \theta(\mu - X_t)e^{\theta t}\right)dt + \sigma e^{\theta t}\,dW_t \\ &= \theta\mu e^{\theta t}\,dt + \sigma e^{\theta t}\,dW_t. \end{aligned}$$
- I.e.:
$$Y_t = Y_0 + \int_0^t \theta\mu e^{\theta u}\,du + \int_0^t \sigma e^{\theta u}\,dW_u = (X_0 - \mu) + \mu e^{\theta t} + \sigma\int_0^t e^{\theta u}\,dW_u\,.$$
- Thus:
$$X_t = (X_0 - \mu)e^{-\theta t} + \mu + \sigma\int_0^t e^{\theta(u-t)}\,dW_u = (X_0 - \mu)e^{-\theta t} + \mu + \sigma\int_{s=0}^t e^{-\theta s}\,dW_{t-s}\,.$$
- Define $Z_t = X_{t+\tau}$ (i.e. Z_t is an Ornstein-Uhlenbeck process started at time $-\tau$). Then, in the limit as $\tau \to \infty$:
$$Z_t = \mu + \sigma\int_{s=0}^{\infty} e^{-\theta s}\,dW_{t-s}\,.$$

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

- 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.
- 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 $e^{-i\phi} = \cos\phi + i \sin\phi$ where $i = \sqrt{-1}$, 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:

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

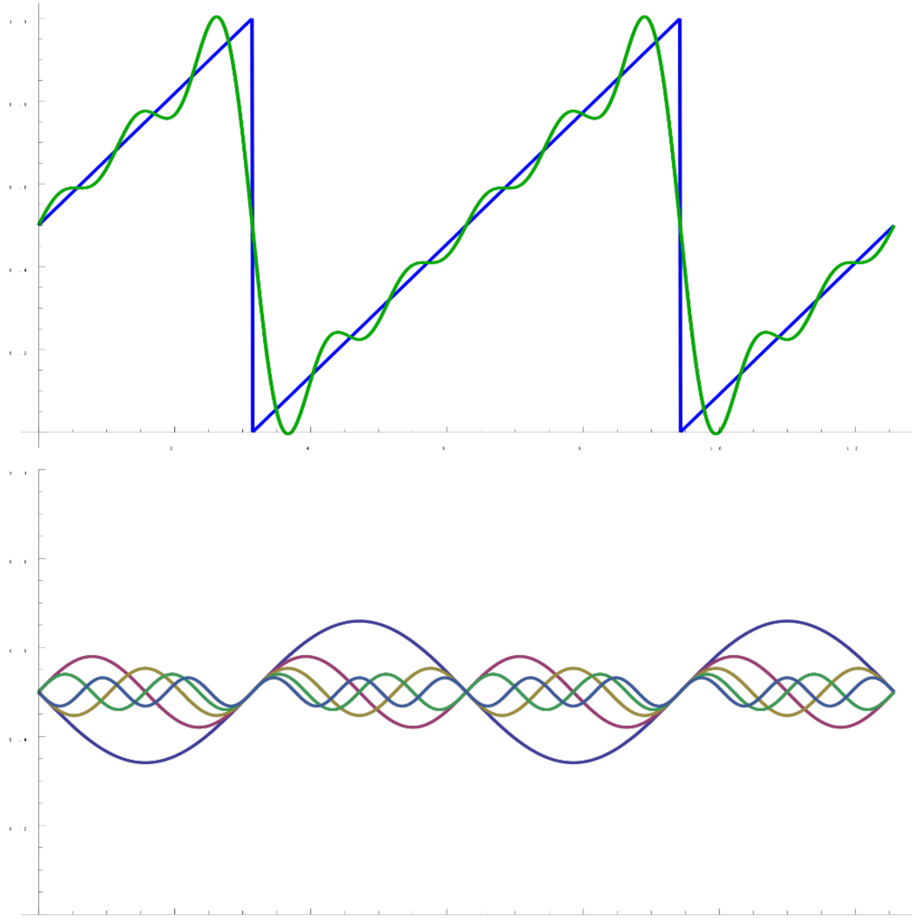
- Then for almost all $x \in [0,1]$:

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

- 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 $\mathcal{F}^{-1} a = \sum_{n=-\infty}^{\infty} a e_n$.

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!
  - Exercise, prove that if $\mathcal{F}(a)$ is viewed as a function on the whole real line, then $\mathcal{F}(a)(\xi) = \mathcal{F}(a)(\xi + 1)$.
- As before, $\mathcal{F}$ is invertible, in the sense that $\left(\mathcal{F}^{-1}\big(\mathcal{F}(f)\big)\right)_n = a_n$ 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 on the real line

- The Fourier transform may also be defined for functions on the real line.
- 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:
$$\mathcal{F}^{-1}(g)(x) = \int_{-\infty}^{\infty} g(\xi) e^{2\pi i x \xi} \, d\xi \, .$$
- When we are working with continuous time stochastic processes, this is the Fourier transform we use.
  - 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 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.

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 $\tau > 0$:

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

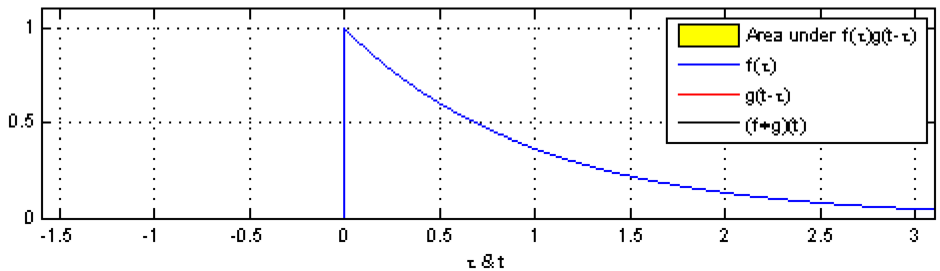
- Then:

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

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:

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

- One characterisation of "white noise" is that it has constant spectral density. I.e. it may be shown that $\mathbb{E}\left|\mathcal{F}(DW)\left(\frac{\omega}{2\pi}\right)\right|^2 = 1$.
- So:

$$S_{ZZ}(\omega) = \mathbb{E}\left|\mathcal{F}(Z_\cdot - \mathbb{E}Z_\cdot)\left(\frac{\omega}{2\pi}\right)\right|^2 = \sigma^2 \left|\int_{-\infty}^{\infty} \mathbb{1}(s > 0)e^{-\theta s}e^{-i\omega s}\, ds\right|^2 = \sigma^2 \left|\frac{1}{\theta + i\omega}\right|^2 = \frac{\sigma^2}{\theta^2 + \omega^2}.$$

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, $S_{\varepsilon\varepsilon}(\omega) = 1$.
  - Exercise: prove this.
- 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 $\Theta_1(\lambda) = 1 + \theta\lambda$:

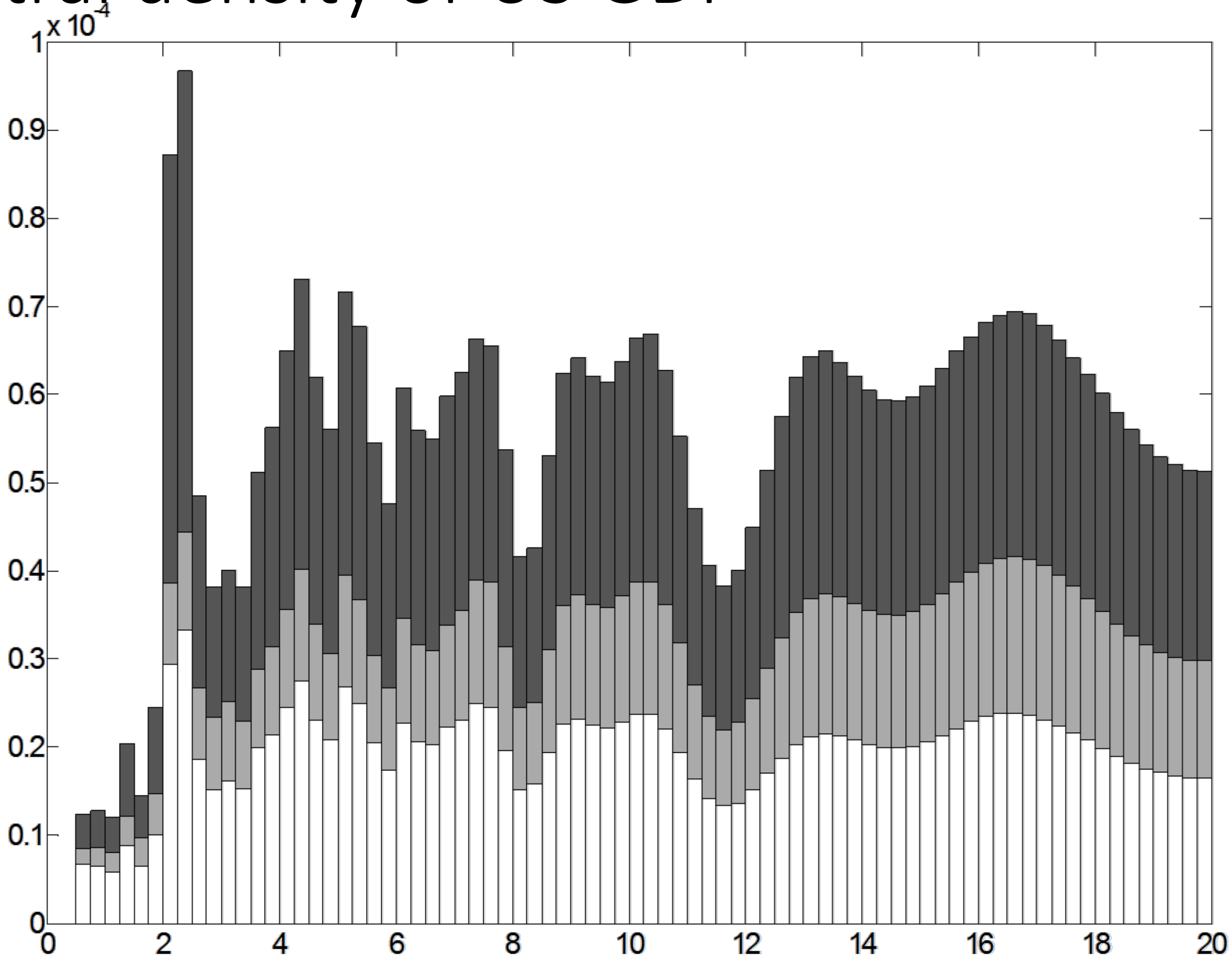
$$S_{yy}(\omega) = \sigma^2 \frac{\left(1 + \theta e^{-i\omega}\right)\left(1 + \theta e^{i\omega}\right)}{(1 - \phi e^{-i\omega})(1 - \phi e^{i\omega})} = \sigma^2 \frac{1 + 2\theta \cos \omega + \theta^2}{1 - 2\phi \cos \omega + \phi^2}.$$

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 last year 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.

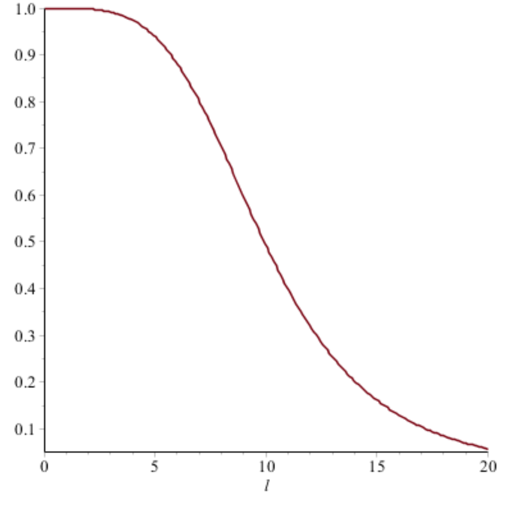
- Problems:
  - Since this is a time domain filter, there's no guarantee it's going to recover the frequencies we're interested in.
  - The filter is non-causal, i.e. the filtered observation at t depends on the source data at $t + 1$, $t + 2$, etc.
  - The filter suffers from "end-point bias", with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

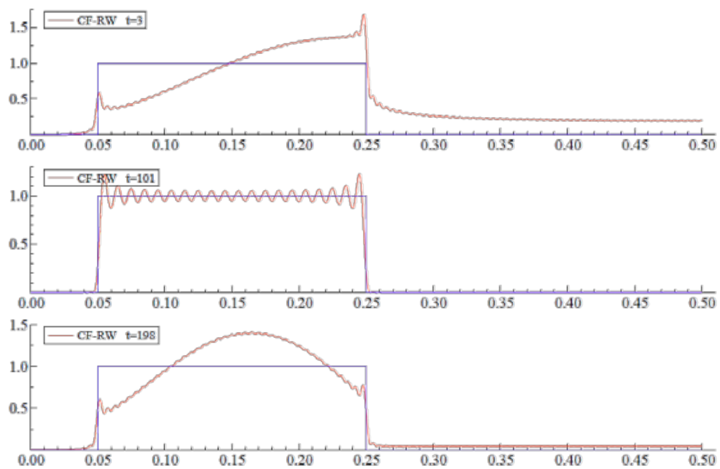


Figure 2.2: Squared gain of ideal and CF-RW filter with $T = 201$ observations at various times t, with pass band [1/20, 1/4] (frequency as number of cycles per period).

Conclusion and recap

- Reduced form VARs do not identify shocks.
- Identification is impossible without making strong prior assumptions.
- Continuous time stochastic processes are not so different to discrete time ones.
- Care must be taken when filtering the data.