Lecture 3: Linear Time Series Methods

Christian Wolf

MIT

14.461, Fall 2022

Today's lecture

- Last time: linearized structural macro models induce SVMA representations

$$y_t = \sum_{\ell=0}^{\infty} \Theta_\ell \varepsilon_{t-\ell}$$

- Next couple of lectures: how can we use time series data to learn about the Θ_ℓ's?
- Today: crash course on **time series fundamentals**
  - Basic concepts: autocovariances, spectra, projections
  - Linear models: VMA, VAR, VARMA
  - Wold decomposition

See syllabus for textbook treatments. This will be a highly selective review, and very far from any research frontier. Much more detailed coverage in 14.384.

Outline

Outline

Time series analysis

- Time series: data with a **time ordering**
  - Series have trends and are correlated over time. We see one realization ("history")
  - Inference requires us to make assumptions ensuring that "the present is like the past", in some loose sense
- Let's start by defining stochastic processes over time:

Definition

An n-dimensional stochastic process is a collection $\{y_t\}_{t \in \mathcal{T}}$ of n-dimensional vectors defined on a probability space $(\Omega, \mathcal{F}, P)$.

- The distribution of a stochastic process is summarized by distribution functions:

$$F_{t_1,\ldots,t_k}(y_1, \ldots, y_k) \equiv P(y_{t_1} \leq y_1, \ldots, y_{t_k} \leq y_k)$$

  for all finite collections of time points $t_1, t_2, \ldots, t_k \in \mathcal{T}$
- Randomness is across different histories of y. We only see one.

Strict stationarity

- What does it mean for the present to be like the past?
- One natural starting point is the assumption of **(strict) stationarity**:

Definition

A stochastic process $\{y_t\}$ $(t = 0, 1, 2, \dots)$ is **strictly stationary** if

$$(y_t, \dots, y_{t+k}) \overset{d}{=} (y_{t+\ell}, \dots, y_{t+\ell+k})$$

for all k, ℓ, where "$\overset{d}{=}$" means "has the same joint distribution as".

  - In words: the distribution of a subsample of any given length does not depend on the point in time at which the subsample starts
  - Given such an assumption (+ "independence for far enough y's"), a realization of a stochastic process could allow us to learn about the distribution function P
    Note: formalization of independence notion is the assumption of ergodicity.

Covariance stationarity

- In this class we pay particular attention to *second-moment properties* of time series:
  One reason: given short aggregate time series, higher-order moments are very hard to estimate.

Definition

A stochastic process $\{y_t\}$ is **weakly (or covariance) stationary** if:

1. $\mathbb{E}(y_t)$ does not depend on t
2. $\text{Cov}(y_t, y_{t-\ell}) \equiv \mathbb{E}\left[(y_t - \mathbb{E}(y_t))(y_{t-\ell} - \mathbb{E}(y_{t-\ell}))'\right]$ exists, is finite, and depends only on ℓ, not t

- In light of this it makes sense to define:
  a) Mean: $\mu_y = \mathbb{E}(y_t)$
  b) Covariance: $\Gamma_y(\ell) \equiv \text{Cov}(y_t, y_{t-\ell})$
- We will throughout be studying **covariance-stationary** processes
  Will use: those second-moment properties—i.e., $\mathbb{E}(\bullet)$ and $\text{Cov}(\bullet)$—are estimable.

Outline

Autocovariance function

- Write the **autocovariance function** as

$$\Gamma_y(\ell) \equiv \begin{pmatrix} \mathrm{Cov}(y_{1t}, y_{1t-\ell}) & \dots & \mathrm{Cov}(y_{1t}, y_{nt-\ell}) \\ \vdots & \ddots & \vdots \\ \mathrm{Cov}(y_{nt}, y_{1t-\ell}) & \dots & \mathrm{Cov}(y_{nt}, y_{nt-\ell}) \end{pmatrix}$$

  I.e., for each ℓ, $\Gamma_y(\ell)$ is an $n \times n$ matrix

- It has the following properties:
  a) $\Gamma_y(\ell) = \Gamma_y(-\ell)'$
  b) $|\Gamma_{y,ij}(\ell)| \leq \sqrt{\Gamma_{y,ii}(0)\Gamma_{y,jj}(0)}$

- The **autocovariance function** is our first of *three fundamental representations*: it fully summarizes all second-moment properties of a covariance-stationary time series process
  As said above, note that this representation is in principle estimable—for a covariance-stationary process we can get all covariances from long enough time series.

Autocorrelation function

- Can similarly define the **autocorrelation function**:

$$R_y(\ell) \equiv \begin{pmatrix} \text{Corr}(y_{1t}, y_{1t-\ell}) & \dots & \text{Corr}(y_{1t}, y_{nt-\ell}) \\ \vdots & \ddots & \vdots \\ \text{Corr}(y_{nt}, y_{1t-\ell}) & \dots & \text{Corr}(y_{nt}, y_{nt-\ell}) \end{pmatrix}$$

  - That is, we have

$$R_{y,ij}(\ell) = \frac{\Gamma_{y,ij}(\ell)}{\sqrt{\Gamma_{y,ii}(0)\Gamma_{y,jj}(0)}}$$

- Same properties, plus $R_{y,ii}(0) = 1$ for all i

Outline

Definition

- The *spectrum* is defined as a function of the autocovariance function

Definition

Let $\{y_t\}$ have an absolutely summable autocovariance function $\Gamma_y(\bullet)$. Then the **spectral density function (or spectrum)** is defined as

$$s_y(\omega) \equiv \frac{1}{2\pi} \sum_{\ell=-\infty}^{\infty} e^{-i\omega\ell}\Gamma_y(\ell), \quad \omega \in [-\pi, \pi]$$

  - Note that we can also invert this mapping to get

$$\Gamma_y(\ell) = \int_{-\pi}^{\pi} e^{i\omega\ell} s_y(\omega) d\omega$$

- So the spectral density is just the **Fourier transform of the ACF**. As the mapping is 1-1, it is our second *fundamental representation*. But why should we care about it?

Interpretation

- To make some progress on interpreting the spectrum, consider the following process:

$$y_t = \sum_{j=0}^{N-1} \left\{ u(\omega_j)\cos(\omega_j t) + v(\omega_j)\sin(\omega_j t) \right\}$$

  - Notation: $N \in \mathbb{N}$, $\omega_j \equiv \frac{2\pi j}{N} - \pi$, $\sigma(\bullet)$ is a function from $[-\pi, \pi]$ to $\mathbb{R}_+$ s.t. $\sigma(\omega) = \sigma(-\omega)$ and $(u(\omega_j), v(\omega_j))_{j=0,1,\dots,N-1}$ are mean-zero, uncorrelated random variables with

$$\mathrm{Var}(u(\omega_j)) = \mathrm{Var}(v(\omega_j)) = \frac{2\pi}{N}\sigma(\omega_j)$$

  - In words: y_t is a sum of N cosine and sine waves with frequency ω_j and random and independent weights u and v

Illustration: sine and cosine waves

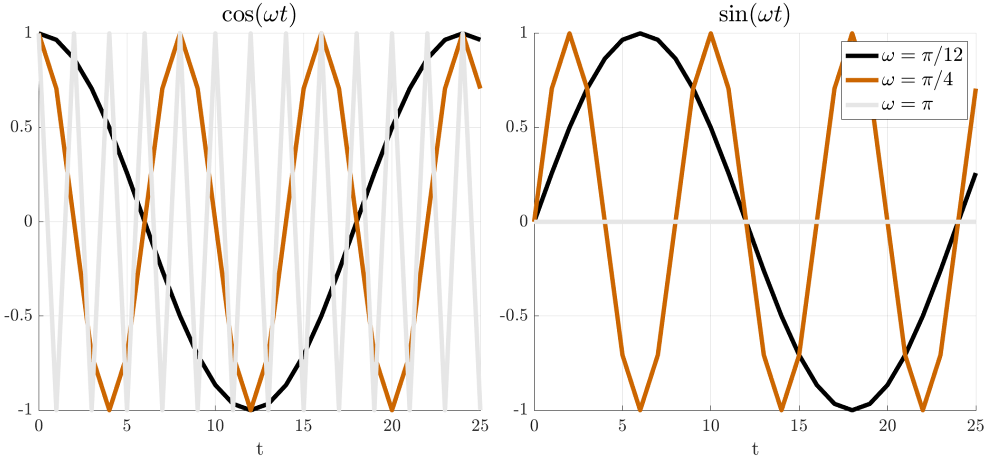
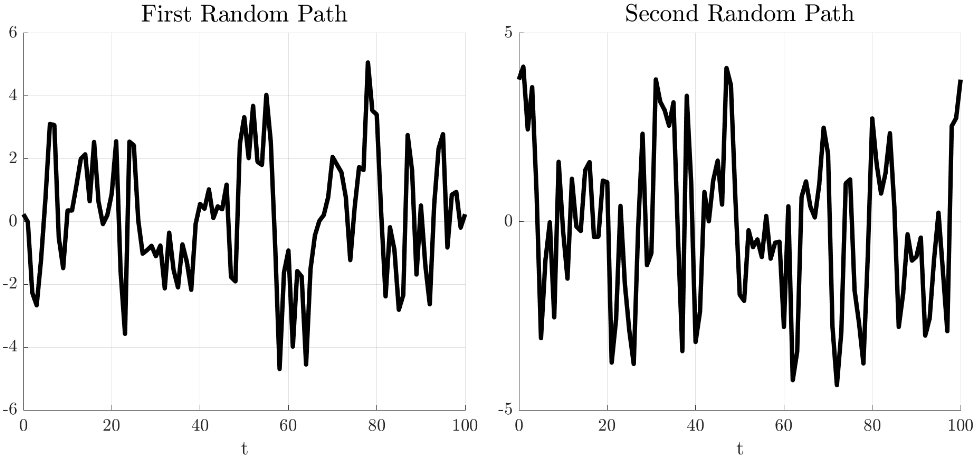


Illustration: random sums of sinusoids



Interpretation

- First observation: y_t is a covariance-stationary process with ACF
  [Verify this! Hint: $\cos(\alpha)\cos(\beta) + \sin(\alpha)\sin(\beta) = \cos(\alpha - \beta)$]

$$\Gamma_y(\ell) = \frac{2\pi}{N} \sum_{j=0}^{N-1} \sigma(\omega_j) \cos(\omega_j \ell)$$

- Second observation: the $\sigma \to$ ACF link looks a lot like the spectral density $\to$ ACF link
  - Note that, for a scalar time series, we have that [Hint: $e^{i\omega\ell} = \cos(\omega\ell) + i\sin(\omega\ell)$ and then use the fact that $s_y(\omega) = s_y(-\omega)$.]

$$\Gamma_y(\ell) = \int_{-\pi}^{\pi} e^{i\omega\ell} s_y(\omega) d\omega = \int_{-\pi}^{\pi} \cos(\omega\ell) s_y(\omega) d\omega$$

    This looks like the previous expression, but with $N \to \infty$ (very loosely)
  - But our function $\sigma(\bullet)$ has a clear interpretation: it is the variance corresponding to cycles with frequencies ω_j. Does the spectral density have a similar interpretation?

Interpretation

- Affirmative answer is provided by the **spectral representation theorem**:
  - We can write *every* covariance-stationary, mean-0 time series as

$$y_t = \int_{-\pi}^{\pi} \cos(\omega t) du(\omega) + \int_{-\pi}^{\pi} \sin(\omega t) dv(\omega) = \int_{-\pi}^{\pi} e^{i\omega t} dz(\omega)$$

    where u and v are orthogonal "innovations" for each ω and $dz(\omega) = du(\omega) - idv(\omega)$
  - Let $\sigma_y(\omega)d\omega = \text{Var}(dz(\omega))$. Can show that indeed $\sigma_y(\omega) = s_y(\omega)$ is the spectrum.
- **Takeaways**:
  1. Spectral representations split time series into independent components. That's useful: replace complicated dependence structures over time with simpler, independent pieces across frequencies.
  2. $s_y(\omega)$ has a clean interpretation as the volatility of each of these independent pieces

Outline

Linear projections

- Many of our arguments will exploit **linear projections**. Easy with finite dimensions:

Definition

Let y be a scalar random variable, and x be an n-dimensional random vector. The best linear predictor of y given x is given by

$$\mathbb{E}^*(y \mid x) \equiv \beta' x$$

where

$$\beta \equiv \operatorname*{argmin}_{b \in \mathbb{R}^n} \mathbb{E}\left[(y - b'x)^2\right]$$

Note that this implicitly requires finite second moments.

- The best linear predictor is often referred to as the linear (or least-squares) projection

Linear projections

- In this class we will routinely project onto ***infinitely many* random variables**, e.g. all current and past values of some macro observables x_t, $\{x_{t-\ell}\}_{\ell=0}^{\infty}$
- Formal way to do so is Hilbert space theory. Sketch of key ideas:
  [see chapter 2 of Brockwell & Davis for the formal treatment]
  - Let $\{x_i\}_{i\in\mathcal{I}}$ be a collection of scalar random variables. Let $\text{span}(x_i, i \in \mathcal{I})$ denote the space of all limits of (weighted) sums of the x_i's.
  - There exists a unique random variable $\hat{y} \in \text{span}(x_i, i \in \mathcal{I})$ such that

$$\mathbb{E}\left[(y-\hat{y})^2\right] = \inf_{z\in\text{span}(x_i, i\in\mathcal{I})} \mathbb{E}\left[(y-z)^2\right]$$

  - $\hat{y} \equiv \mathbb{E}^*(y \mid \{x_i\}_{i\in\mathcal{I}})$ is the best linear prediction. It satisfies

$$\mathbb{E}\left[(y-\hat{y})\,\tilde{x}\right] = 0 \quad \text{for all } \tilde{x} \in \text{span}(x_i, i \in \mathcal{I})$$

Outline

White Noise

- Our key building block process will be white noise:

Definition

An n-dimensional covariance-stationary process $\{y_t\}$ is **white noise** if $\mu_y = 0$, $\Gamma_y(0) = \Sigma$ and $\Gamma_y(\ell) = 0$ for all $\ell \neq 0$. We write $y_t \sim WN(0, \Sigma)$.

- Note that a white noise is *linearly* unpredictable based on its own lags:

$$\mathbb{E}^* \left(y_{it} \mid \{y_\tau\}_{-\infty<\tau<t} \right) = \mathbb{E}(y_{it}) = 0$$

- A white noise process may however be nonlinearly predictable
  - Example: $y_t = \cos(\omega t)$ where $\omega \sim \text{uniform}[0, 2\pi]$.
    Hint: to show white noise property use $\cos(t_1\omega)\cos(t_2\omega) = \frac{1}{2}(\cos((t_1 + t_2)\omega) + \cos((t_1 - t_2)\omega))$.

Lag operators

- A useful object in time series analysis are so-called **lag operators**
- If $\{y_t\}$ is a stochastic process, then the lag operator L is defined such that

$$Ly_t = y_{t-1}, \quad \text{for all } t$$

- Some properties:
  - L is a linear operator
  - L^{-1} exists and is given by $L^{-1}y_t = y_{t+1}$. It is also called the **lead operator**
  - For any $d \in \mathbb{Z}$, we have $L^d = L(L(\ldots(Ly_t)\ldots)) = y_{t-d}$

Lag polynomials

- Using **lag operators** we can define **lag polynomials**
  - Let $\Psi(z) = \sum_{\ell=-\infty}^{\infty} \Psi_\ell z^\ell$ denote a matrix polynomial in the scalar z, and suppose the Ψ_ℓ's are absolutely summable across ℓ. Define the lag polynomial

$$\Psi(L) \equiv \sum_{\ell=-\infty}^{\infty} \Psi_\ell L^\ell$$

  - Given the definition of lag operators, applying a lag polynomial to a stochastic process $\{y_t\}$ simply means that

$$\Psi(L) y_t = \sum_{\ell=-\infty}^{\infty} \Psi_\ell y_{t-\ell}$$

- Lag polynomials can be either **two-sided** or **one-sided**:
  - Two-sided: $\Psi(L) \equiv \sum_{\ell=-\infty}^{\infty} \Psi_\ell L^\ell$ = looks into the past & future
  - One-sided: $\Psi(L) \equiv \sum_{\ell=0}^{\infty} \Psi_\ell L^\ell$ = only looks into the past

Lag polynomials

- Some **properties** of lag polynomials:
  - We can combine conformable lag polynomials, e.g.

$$\zeta(L) \equiv \Psi(L)\Lambda(L) = \sum_{\ell=-\infty}^{\infty} \zeta_\ell L^\ell, \quad \text{where } \zeta_\ell = \sum_{m=-\infty}^{\infty} \Psi_m \Lambda_{\ell-m}$$

  - If c is a constant vector, then $\Psi(L)c = \Psi(1)c = (\sum_{\ell=-\infty}^{\infty} \Psi_\ell)c$
- Using white noise and lag operators, we can finally define the **kinds of time series processes** that we will study in this class ...

Outline

Vector moving average

- A key process for us will be **vector moving averages**:

Definition

Let $\mu \in \mathbb{R}^n$, let Σ and Θ_ℓ ($\ell = 1, 2, \ldots, q$) be $n \times n$ matrices, and set $\Theta_0 = I$. The process

$$y_t = \mu + \sum_{\ell=0}^{q} \Theta_\ell z_{t-\ell} = \mu + \Theta(L) z_t, \quad z_t \sim WN(0, \Sigma)$$

is called a **vector moving average** of order q, VMA(q).

  - VMAs are simply linear combinations of white noise processes

- We will pay particular attention to VMA(∞) processes. As we have seen, essentially all linearized structural macro models admit a *structural* VMA(∞) representation.

Vector moving average: properties

Straightforward to arrive at the VMA's second-moment properties:

1. Impulse responses:

$$\mathbb{E}^*(y_{t+\ell} - \mu_y \mid z_t) = \Theta_\ell z_t$$

2. Autocovariance function:

$$\Gamma_y(\ell) = \begin{cases} \sum_{m=0}^{q-\ell} \Theta_m \Sigma \Theta'_{m+\ell} & \text{if } 0 \leq \ell \leq q \\ 0 & \text{otherwise} \end{cases}$$

3. Spectrum:

$$s_y(\omega) = \frac{1}{2\pi} \sum_{\ell=-\infty}^{\infty} e^{-i\omega\ell} \Gamma_y(\ell)$$

Filters

- Could also take linear combinations of more **general time series processes**
- That's what so-called **filters** do:
  - A filter takes linear combos of a time series process to map that process into a new one:

$$x_t \equiv \Psi(L) y_t = \sum_{\ell=\infty}^{\infty} \Psi_\ell y_{t-\ell}$$

A simple example would be the first difference: $x_t \equiv y_t - y_{t-1}$

  - Verify that the ACF of a filtered series is given as

$$\Gamma_x(\ell) = \sum_{k=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \Psi_k \Gamma_y(\ell + m - k) \Psi'_m$$

  - For the spectral density it can be shown that: [scalar case for simplicity]

$$s_x(\omega) = |\Psi(e^{-i\omega})|^2 s_y(\omega) = |\sum_{\ell=-\infty}^{\infty} \Psi_\ell e^{-i\omega\ell}|^2 s_y(\omega)$$

Example: band-pass filter

- Filters are useful to isolate **fluctuations at certain frequencies** = "band-pass filter"

$$\Psi(e^{-i\omega}) = \begin{cases} 1 & \text{if } |\omega| \in [\alpha, \beta] \\ 0 & \text{otherwise} \end{cases}$$

  - The resulting series consists only of the sine and cosine waves at frequencies in $[\alpha, \beta]$
  - The total variance thus for example only reflects volatility at those frequencies:

$$\text{Var}(x_t) = \text{Var}(\Psi(L)y_t) = \frac{1}{2\pi}\int_{-\pi}^{\pi} |\Psi(e^{-i\omega}|^2 s_y(\omega)d\omega = \frac{1}{2\pi}\int_{|\omega|\in[\alpha,\beta]} s_y(\omega)d\omega$$

- Using filters that isolate frequencies between 2 and 32 quarters tends to give something reasonably similar to the well-known **Hodrick-Prescott filter**

Example: band-pass filter

- Note that the band-pass filter is a two-sided filter:

$$\Psi_\ell = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i\omega\ell}\Psi(e^{-i\omega})d\omega = \frac{1}{2\pi}\int_{|\omega|\in[\alpha,\beta]} e^{i\omega\ell}d\omega = \frac{\sin(\ell\beta)}{\pi\ell} - \frac{\sin(\ell\alpha)}{\pi\ell}$$

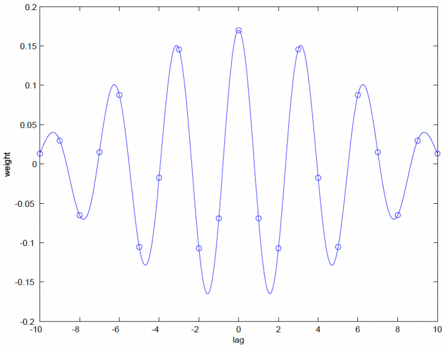


Figure 8.7: Moving average weights of a bandpass filter.

Source: Cochrane (2005)

Vector autoregression

- The second key process will be **vector autoregressions**:

Definition

Let $\nu \in \mathbb{R}^n$, let Σ and A_ℓ ($\ell = 1, 2, \ldots, p$) be $n \times n$ matrices. A covariance-stationary process satisfying

$$y_t = \nu + \sum_{\ell=1}^{p} A_\ell y_{t-\ell} + z_t, \quad z_t \sim WN(0, \Sigma)$$

is called a **Vector Autoregression** of order p, VAR(p).

- Note that, with the matrix polynomial $A(L) \equiv I - \sum_{\ell=1}^{p} A_\ell L^\ell$, we may write this as

$$A(L) y_t = \nu + z_t$$

Vector autoregression: stationarity

- Note that **stationarity is not guaranteed**; i.e., a covariance-stationary process satisfying the VAR equation may not exist. When does it exist?
- Sufficient condition: existence of a (one-sided) **inverse** of $A(L)$
  - A lag polynomial $\Psi(L)$ is called a one-sided inverse of $A(L)$ if $\Psi(L) = \sum_{\ell=0}^{\infty} \Psi_\ell L^\ell$ is absolutely summable and
    $$\Psi(L)A(L) = I$$
    We write $\Psi(L) = A(L)^{-1}$
  - We thus get
    $$y_t = \Psi(L)A(L)y_t = \mu + \Psi(L)z_t, \quad \mu \equiv \Psi(1)\nu$$
    so we have mapped the VAR(p) into a VMA(∞).
- When does $A(L)^{-1}$ exist? need all roots of $\det(A(z))$ to be outside the unit circle
  See Brockwell-Davis, Th'm 11.3.1 for the full result. For intuition, consider an AR(1), $y_t = \rho y_{t-1} + z_t$. Can solve out past y's if $\rho \in (-1, 1)$.

Vector autoregression: properties

Slightly more involved to arrive at second-moment properties:

1. Impulse responses are given via $\Psi(L)$:

$$\Psi_0 = I, \quad \Psi_\ell = \sum_{m=1}^{\min(\ell,p)} A_m \Psi_{\ell-m}$$

2. Autocovariance function:

$$\Gamma_y(\ell) = \begin{cases} \sum_{m=1}^{p} A_m \Gamma_y(m)' + \Sigma & \text{if } \ell = 0 \\ \sum_{m=1}^{p} A_m \Gamma_y(\ell - m) & \ell \geq 1 \end{cases}$$

3. Spectrum:

$$s_y(\omega) = \frac{1}{2\pi} \sum_{\ell=-\infty}^{\infty} e^{-i\omega\ell} \Gamma_y(\ell)$$

VARMA model

- Finally we can combine VMAs and VARs to obtain **VARMAs**:

Definition

Let $\mu \in \mathbb{R}^n$, and Σ, A_ℓ $(\ell = 1, 2, \ldots, p)$ and Θ_ℓ $(\ell = 1, 2, \ldots, q)$ be $n \times n$ matrices. A covariance-stationary process satisfying

$$A(L)(y_t - \mu) = \Theta(L)z_t, \quad z_t \sim WN(0, \Sigma)$$

is called a VARMA(p,q) process.

- Stationarity properties as well as expressions for impulse responses, autocovariances and spectra generalize straightforwardly from the VMA and VAR cases

Causality & Invertibility

Two important properties of VARMA processes are **causality** and **invertibility**:

Definition

A VARMA process $\{y_t\}$ is said to be **causal** with respect to $\{z_t\}$ if

$$y_t \in \text{span}(z_\tau, -\infty < \tau \leq t)$$

- In words: can write y_t as function of current and lagged white noise realizations, $z_{t-\ell}$
- Sufficient condition is that $A(L)$ has 1-sided inverse, giving VMA(∞) representation

$\Rightarrow$ Our structural macro models yield VMA representations and so in particular always give causal VARMA processes for the observables y_t

Causality & Invertibility

Two important properties of VARMA processes are **causality** and **invertibility**:

Definition

A VARMA process $\{y_t\}$ is said to be **invertible** with respect to $\{z_t\}$ if

$$z_t \in \text{span}(y_\tau, -\infty < \tau \le t)$$

- Can obtain the white noise realizations z_t as a function of current and lagged values of the process itself, $y_{t-\ell}$.
- Sufficient condition is that $\Theta(L)$ has 1-sided inverse, giving a VAR(∞) representation
  - $\Rightarrow$ This property is far from guaranteed in our structural models. E.g. if we have 5 shocks in z but only 2 observables in y, then the process can't possibly be invertible.

Illustrating invertibility

- Let's provide a quick illustration of invertibility using the canonical three-equation NK model Galí (2007), Woodford (2011)
  - Consider a model with three shocks and no endogenous persistence:

$$
\begin{aligned}
y_t &= \mathbb{E}_t(y_{t+1}) - (i_t - \mathbb{E}_t(\pi_{t+1})) + \sigma^d \varepsilon_t^d && \text{(IS)} \\
\pi_t &= \kappa y_t + \beta \mathbb{E}_t(\pi_{t+1}) - \sigma^s \varepsilon_t^s && \text{(NKPC)} \\
i_t &= \phi_\pi \pi_t + \sigma^m \varepsilon_t^m && \text{(TR)}
\end{aligned}
$$

  - Solving the model gives a static VMA representation:

$$
\begin{pmatrix} y_t \\ \pi_t \\ i_t \end{pmatrix} = \underbrace{\frac{1}{1+\phi_\pi \kappa} \begin{pmatrix} \sigma^d & \phi_\pi \sigma^s & -\sigma^m \\ \kappa \sigma^d & -\sigma^s & -\kappa \sigma^m \\ \phi_\pi \kappa \sigma^d & -\phi_\pi \sigma^s & \sigma^m \end{pmatrix}}_{\Theta} \times \begin{pmatrix} \varepsilon_t^d \\ \varepsilon_t^s \\ \varepsilon_t^m \end{pmatrix}
$$

Illustrating invertibility

$$\begin{pmatrix} y_t \\ \pi_t \\ i_t \end{pmatrix} = \underbrace{\frac{1}{1+\phi_\pi \kappa} \begin{pmatrix} \sigma^d & \phi_\pi \sigma^s & -\sigma^m \\ \kappa \sigma^d & -\sigma^s & -\kappa \sigma^m \\ \phi_\pi \kappa \sigma^d & -\phi_\pi \sigma^s & \sigma^m \end{pmatrix}}_{\Theta} \times \begin{pmatrix} \varepsilon_t^d \\ \varepsilon_t^s \\ \varepsilon_t^m \end{pmatrix}$$

- It's easy to verify that, for standard parameter values, Θ is invertible
- But now suppose that we only observe y_t and π_t:
  - Then we certainly can't disentangle all three shocks $\{\varepsilon_t^d, \varepsilon_t^s, \varepsilon_t^m\}$
  - More precisely: can back out ε_t^s, but impossible to disentangle ε_t^d and ε_t^m
    Can you see why?

Outline

Wold decomposition

Proposition

Let $\{y_t\}$ be an n-dimensional covariance-stationary time series. There exists an $n \times n$ lag polynomial $\Psi(L)$ and n-dimensional $u_t \sim WN(0, \Sigma)$ such that

$$y_t = \Psi(L)u_t + d_t, \quad \Psi(L) = I + \sum_{\ell=1}^{\infty} \Psi_\ell L^\ell$$

and where:

1. $\Psi(L)$ *is square-summable*
2. $u_t = y_t - \mathbb{E}^*(y_t \mid \{y_\tau\}_{-\infty<\tau\leq t-1})$, $\mathbb{E}^*(u_t \mid \{y_\tau\}_{-\infty<\tau\leq t-1}) = 0$, $\Sigma = var^*(y_t \mid \{y_\tau\}_{-\infty<\tau\leq t-1})$
3. $\{y_t\}$ *is invertible with respect to* u_t, *i.e.* $u_t \in \operatorname{span}(y_\tau, -\infty < \tau \leq t)$
4. $\{d_t\}$ *is a purely deterministic process, i.e.* $var^*(d_t \mid \{d_\tau\}_{-\infty<\tau\leq t-1}) = 0$

Wold decomposition: discussion

- The decomposition says that *any* covariance-stationary time series can be written as

$$\textbf{VMA}(\infty) + \textbf{deterministic component}$$

  - Note: this formally justifies the loose impulse-propagation discussion in Lecture 1

- Interpretation of the u's
  - The Wold decomposition does something very simple: it splits a process $\{y_t\}$ into one-step-ahead prediction errors and a perfectly predictable residual
  - Note: we can thus also turn the Wold decomposition into a VAR(∞)

$$A(L)y_t = u_t + \tilde{d}_t, \quad A(L) = \Psi(L)^{-1}, \quad \tilde{d}_t = \Psi(1)^{-1} d_t$$

Wold decomposition: discussion

- The Wold decomposition is nothing but yet another way of summarizing the **second-moment properties** of a time series process
  - Nothing guarantees that the Ψ's are interesting. They are just coefficients on reduced-form prediction errors.
  - We can freely map between autocovariance functions and the Wold decomposition:

$$\Psi_\ell = \mathrm{Cov}(y_t, u_{t-\ell})\Sigma^{-1}$$

  - The Wold decomposition is our *third fundamental representation*. For second-order properties, we can freely map between ACF, spectrum, and Wold decomposition.
- It thus follows in particular that the Wold decomposition is **identifiable** from aggregate time series data (just like autocovariances & spectral densities)

Summary

- We saw some **basic time series models/concepts**
- So far everything was **reduced-form**:
  - Presented ACF/spectral density/Wold decomposition as three ways of summarizing the second-moment properties of observable time series data
  - These reduced-form objects are in principle estimable, but of course nothing says that they are *interesting*, i.e. related to our Θ's in *structural* VMA representations
- Next: what kind of **addt'l economic assumptions** allow us to learn about the Θ's?