Lecture 2: linear time series methods

Diego Känzig
Northwestern University

416-2 Empirical Macro, Winter 2025

Today's lecture

- Last time, we introduced the impulse-propagation/Slutzky-Frisch paradigm

$$y_t = \sum_{l=0}^{\infty} \Theta_l \varepsilon_{t-l}$$

- Today: Why this is a good framework to think about economic fluctuations?
- Next couple lectures: How can we use time series data to learn about the Θ_l's
- But first: refresher on **time series fundamentals**
- The exposition here builds on the excellent materials by Christian Wolf: `https://www.christiankwolf.com/teaching`
- For textbook treatment, see
  - Hamilton 1994 (classic reference) or
  - Brockwell and Davis 2009 (more technical)

Outline

Time series analysis

- Time series: data with a **time ordering**

Definition (Time series)

$\{y_t\}_{t \in \mathcal{T}}$: Set of observations y_t, recorded at a specified time $t \in \mathcal{T}$

- Series can have trends and may be correlated over time
  - How to deal with that?
- For valid inference required to make assumptions ensuring that "the present is like the past" (at least in some loose sense)

Time series analysis

Definition (Stochastic process)

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 $t_1, \ldots, t_k \in \mathcal{T}$
- Time series is a realization of a stochastic process
- Randomness is across different histories of y. But we only see one!

Strict stationarity

- For inference, we need to make assumptions on how the past links to the present
    - What does it mean for the present to be like the past?
- One natural starting point is the assumption of **(strict) stationarity**:

Definition (Strict stationarity)

A stochastic process $\{y_t\}_{t \in \mathcal{T}}$ is **strictly stationary** if $(y_t, \ldots, y_{t+k})$ has the same joint distribution as $(y_{t+l}, \ldots, y_{t+l+k})$ for all k, l.

- **In words**: the distribution of a subsample of any given length does not depend on the point in time at which the subsample starts
- Such an assumption combined with some notion of "independence for far enough y's" allows us to learn about the distribution function P

Covariance stationarity

- In this class we focus on **second-moment properties** of time series
  - One reason: given short time series, higher-order moments very hard to estimate

Definition (Weak stationarity)

A stochastic process $\{y_t\}$ is **weakly (or covariance) stationary** if the first and second moments do not depend on t and are finite. That is, for all t,

$$\mathbb{E}[y_t] = \mu < \infty$$

$$Var(y_t) < \infty$$

$$Cov(y_t, y_{t+k}) = Cov(y_{t+l}, y_{t+l+k})$$

- In light of this it makes sense to define:
  - Mean: $\mu_y = \mathbb{E}[y_t]$
  - Covariance: $\Gamma_y(k) = Cov(y_t, y_{t+k})$

Ergodicity

- Formalization of independence notion is the assumption of **ergodicity**

Definition (Ergodicity)

A stationary process $\{y_t\}$ is said to be **ergodic** if for any two bounded functions f and g:

$$\lim_{l\to\infty} |\mathbb{E}[f(y_t,\dots,y_{t+k})g(y_{t+l},\dots,y_{t+l+k})]| \\ = |\mathbb{E}[f(y_t,\dots,y_{t+k})]||\mathbb{E}[g(y_{t+l},\dots,y_{t+l+k})]|$$

- **In words**: ergodicity says that if two sequences of y are “far enough” apart, then one can treat them as independent
- If we have one long history, ergodicity and stationarity mean we can make inference (LLN, CLT) about other potential histories

Do we need stationarity?

- Stationarity facilitates inference but is **not always needed**
- Many objects of interest (IRFs, FEVDs) can be **consistently** estimated even in a non-stationary model
- Classic reference is Sims, Stock, and Watson (1990)
- See also the discussion in Hamilton (1994) and Kilian and Lütkepohl (2017) (Section 3.2.3)

Outline

Three fundamental representations of time series

- Write the **autocovariance function** as

$$\Gamma_y(l) = \begin{pmatrix} \mathsf{Cov}(y_{1,t}, y_{1,t-l}) & \dots & \mathsf{Cov}(y_{1,t}, y_{n,t-l}) \\ \vdots & \ddots & \vdots \\ \mathsf{Cov}(y_{n,t}, y_{1,t-l}) & \dots & \mathsf{Cov}(y_{n,t}, y_{n,t-l}) \end{pmatrix}$$

  i.e. for each k, $\Gamma_y(l)$ is an $n \times n$ matrix
- It has the following properties:
  1. $\Gamma_y(l) = \Gamma_y(-l)'$
  2. $|\Gamma_{y,ij}(l)| \leq \sqrt{\Gamma_{y,ii}(0)\Gamma_{y,jj}(0)}$
- The autocovariance function is our first of *three fundamental representations*: it fully summarizes all second-moment properties of a time series process

Autocorrelation function

- Can similarly define the **autocorrelation function** (ACF)

$$R_y(l) = \begin{pmatrix} \text{Corr}(y_{1,t}, y_{1,t-l}) & \dots & \text{Corr}(y_{1,t}, y_{n,t-l}) \\ \vdots & \ddots & \vdots \\ \text{Corr}(y_{n,t}, y_{1,t-l}) & \dots & \text{Corr}(y_{n,t}, y_{n,t-l}) \end{pmatrix}$$

- We have

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

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

Outline

1. Basic time series concepts
2. Three fundamental time-series representations
   - Autocovariance function
   - Spectrum
   - Additional time-series concepts
   - Wold decomposition
3. Summary

Spectrum

- Our second fundamental representation is the **spectrum**

Theorem (Spectral representation theorem)

The spectral representation theorem states that we can write every covariance-stationary, zero-mean 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 $dz(\omega) = \frac{1}{2}(du(\omega + idv(\omega))$

- The spectrum is defined as $s_y(\omega) = \text{Var}(dz(\omega))$

Spectrum

- **In words**: can represent stationary series as sum of elementary orthogonal period processes, each identified by a given period and multiplied by a random amplitude
- This transformation in the frequency domain allows us to *decorrelate the process*:
    - replace complicated dependence structures over time with simpler, independent pieces across frequencies
    - can study trends, but also cycles of different length and their contribution to the variance of the process
- $s_y(\omega)$ has the **clean interpretation** as the volatility of each of these independent pieces

Relation to the autocovariance function

Definition (Spectral density)

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

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

- Note that we can also invert this mapping to get $\Gamma_y(k) = \int_{-\pi}^{\pi} e^{i\omega k} s_y(\omega) d\omega$
- Spectral density is just the **Fourier transform** of the autocovariance function
- Spectral density function conveys the *same information* as the covariance function

Outline

Additional concepts

- For our last fundamental representation of time series processes, we need to introduce some additional concepts
- Many of our arguments will exploit **linear projections**. Easy with finite dimensions:

Definition (Linear projection)

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

$$\text{Proj}(y|x) \equiv \beta' x, \qquad \text{where } \beta \equiv \text{argmin}_{b \in \mathbb{R}^n} \mathbb{E}[(y - b'x)^2]$$

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 routinely project onto **infinitely many** random variables, e.g. all current and past values of some macro observables x_t, $\{x_{t-l}\}_{l=0}^{\infty}$
- Formal way to do so is Hilbert space theory and the **projection theorem**. Sketch of key ideas:
    - 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 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

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

    - $\hat{y} = \text{Proj}(y | \{x_i\}_{i \in \mathcal{I}})$ is the best linear prediction. It satisfies

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

    - See Brockwell and Davis (2009) for more info

White noise

- A key building block in time series analysis is white noise:

Definition (White noise)

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

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

$$\text{Proj}(y_{i,t} | \{y_\tau\}_{-\infty<\tau<t}) = \mathbb{E}[y_{i,t}] = 0$$

- A white noise process may however be nonlinearly predictable
  - Can you think of an example?

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 $L^{-1}y_t = y_{t+1}$. This is also called the lead operator $F = L^{-1}$
  - 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 $\Psi(z) = \sum_{l=-\infty}^{\infty} \Psi_l z^l$ denote a matrix polynomial in the scalar z, and suppose the Ψ_l's are absolutely summable across l ($\sum_{l=-\infty}^{\infty} |\Psi_l| < \infty$). Define the lag polynomial

$$\Psi(L) = \sum_{l=-\infty}^{\infty} \Psi_l L^l$$

  - Given the definition of lag operator, applying the lag polynomial to a stochastic process $\{y_t\}$ yields

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

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

Lag polynomials

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

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

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

Vector moving average

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

Definition (Vector moving average)

Let Σ and Θ_l ($l = 1, 2, \ldots, q$) be $n \times n$ matrices and set $\Theta_0 = I$. The process

$$y_t = \sum_{l=0}^{q} \Theta_l z_{t-l} = \Theta(L) z_t, \quad z_t \sim WN(0, \Sigma)$$

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**. Why?

Most macro models have VMA representation

- Structural business-cycle models (RBC, NK, HANK, ...) are mappings from **structural shocks** ε_t to **macroeconomic aggregates** y_t
- Typically solving these models includes **two steps**
  1. Linearizing the model's equilibrium conditions to arrive at the form

$$\Gamma_0 \hat{x}_t = \Gamma_1 \hat{x}_{t-1} + \Psi \varepsilon_t + \Pi \eta_t \tag{1}$$

     - $\hat{x}_t$ contains all model variables in (log-)deviation from the deterministic steady state
     - $\varepsilon_t \sim N(0, I)$ are structural shocks. Orthogonal by assumption, unit variance is normalization, normality for convenience
     - η_t is a vector of expectational errors satisfying $\mathbb{E}[\eta_{t+1}] = 0$, indicating which of the equations in (1) hold only in expectation
  2. Solve (1) to obtain a mapping from **shocks** to **macro variables** in state-space form:

$$\hat{x}_t = A\hat{x}_{t-1} + B\varepsilon_t \qquad \text{(State equation)}$$

$$\hat{y}_t = C\hat{x}_t + D\varepsilon_t \qquad \text{(Observation equation)}$$

Most macro models have VMA representation

- Substituting recursively, using stability of the system yields:

$$\hat{y}_t = D\varepsilon_t + CB\varepsilon_{t-1} + CAB\varepsilon_{t-2} + CA^2B\varepsilon_{t-3} + \ldots \equiv \sum_{l=0}^{\infty} \Theta_l \varepsilon_{t-l}$$

- This is a **structural VMA(∞) representation**: mapping the history of shocks ε_t to y_t via the Θ's

Digression: Sequence-space representation

- An alternative way to represent linear RE models is in **sequence space**
- We let boldface denote sequences, e.g.

$$\boldsymbol{y} = (y_0, y_1, y_2, \ldots)'$$

  Here, y is the perfect-foresight transition path from $t = 0$ to ∞ given exogenous shock paths ε

- Let x_t denote a model's endogenous variables and ε_t its shocks. A **perfect-foresight equilibrium** given shock paths $\boldsymbol{\varepsilon}$ is a set of paths $\boldsymbol{x}$ such that

$$F(\boldsymbol{x}, \boldsymbol{\varepsilon}) = 0$$

  where $F(\bullet)$ embeds the model's equilibrium relations (Euler equation, output market-clearing, ...)

Digression: Sequence-space representation

- To first order we can write this as

$$F_x \hat{\boldsymbol{x}} + F_\varepsilon \boldsymbol{\varepsilon} = 0$$

  This defines a mapping from $\boldsymbol{\varepsilon}$'s to $\boldsymbol{x}$'s, just as before

- The solution for our observables $\boldsymbol{y}$ is

$$\hat{\boldsymbol{y}} = \boldsymbol{\Theta}\boldsymbol{\varepsilon}$$

- Note that this solution is immediately in the form of **shock IRFs**
  - The perfect foresight solution has thus directly given us SVMA coefficients Θ_l ...
  - ...which in turn gives the SVMA representation $y_t = \sum_{l=0}^{\infty} \Theta_l \varepsilon_{t-l}$

Sequence- vs. state-space

- Why does considering stochastic shocks in state space and MIT shocks in sequence space give the same answer?

Sequence- vs. state-space

- Why does considering stochastic shocks in state space and MIT shocks in sequence space give the same answer?
- **Intuition**: linearity implies certainty equivalence = perfect foresight (Auclert et al. 2021)
- **Formally**: the Θ's will be the same as before, as we are solving the exact same equations:
  $\rightarrow$ IRFs in stochastic model $= \text{shock}(1, 0, 0, \ldots)' +$ linearized optimality conditions that hold at $t = 0$ and (in expectation) at $t = 1, 2, \ldots +$ return to steady state (stability)

  $\rightarrow$ But the sequence-space approach imposes the exact same linear relations at $t = 0, 1, \ldots +$ return to steady state, so you get the same numbers!

Vector moving average: properties

Back to time series: let's discuss properties of VMAs

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

1. Impulse responses: Proj$(y_{t+l}|z_t) = \Theta_l z_t$
2. Autocovariance function:

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

3. Spectrum:

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

Digression: filters

- We can also take linear combinations of more general time series processes
- That's what **filters** do
  - A filter takes linear combinations of a time series process to map it into a new one:

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

    A simple example would be the first difference: $x_t \equiv y_t - y_{t-1}$ where $\Psi(L) = 1 - L$
  - Verify that the autocovariance function of a filtered series is given as

$$\Gamma_x(l) = \sum_{k=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \Psi_k \Gamma_y(l+m-k) \Psi_m'$$

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

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

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:

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

- A filter that isolates frequencies between 2 and 32 quarters tends to give something reasonably similar to the well-known **Hodrick-Prescott** filter
  - Derive the Hodrick-Prescott filter as an exercise (typically done in the time domain)

Vector autoregression

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

Definition (Vector autoregression)

Let Σ and A_l $(l = 1, 2, \ldots, p)$ be $n \times n$ matrices. A covariance-stationary process satisfying

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

is called a **vector autoregression** of order p, VAR(p)

- With the lag polynomial $A(L) = I - \sum_{l=1}^{p} A_l L^l$ we can also write this as

$$A(L) y_t = z_t$$

Vector autoregression: stationarity

- Note that **stationarity is not guaranteed**, i.e. a covariance-stationary process satisfying the VAR equation may not exist.
- Sufficient condition for stationarity: existence of a (one-sided) inverse of $A(L)$
  - A lag polynomial $\Psi(L)$ is called a one-sided **inverse** of $A(L)$ if $\Psi(L) = \sum_{l=0}^{\infty} \Psi_l L^l$ is absolutely summable and
    $$\Psi(L)A(L) = I$$
    We write $\Psi(L) = A(L)^{-1}$
  - We thus get $y_t = \Psi(L)A(L)y_t = \Psi(L)z_t$, mapping 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, Theorem 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 response functions are given via $\Psi(L)$:

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

2. Autocovariance function:

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

3. Spectrum:

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

VARMA

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

Definition (VARMA)

Let Σ, A_l $(l = 1, 2, \ldots, p)$ and Ω_l $(l = 1, 2, \ldots, q)$ be $n \times n$ matrices. A covariance-stationary process satisfying

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

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 (Causality)

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 WN realizations, z_{t-l}
  - this is a statistical concept and has nothing to do with economic causality
- Sufficient condition: $A(L)$ has one-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 (Invertibility)

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

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

- In words: can obtain WN realizations z_t as function of current and lagged values of process itself, y_{t-k}
- Sufficient condition: $\Theta(L)$ has one-sided inverse, giving 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

Outline

Wold decomposition

Equipped with these tools we can now introduce the last fundamental representation of time series: the **Wold (1954) representation theorem**

Proposition (The Wold decomposition)

Any n-dimensional covariance-stationary time series $\{y_t\}$ *can be written as*

$$y_t = \Psi(L)u_t + d_t = \sum_{l=0}^{\infty} \Psi_l u_{t-l} + d_t, \qquad \textit{where}$$

- $u_t = y_t - \textit{Proj}(y_t | \{y_\tau\}_{-\infty<\tau\leq t-1})$
  with $\textit{Proj}(u_t | \{y_\tau\}_{-\infty<\tau\leq t-1}) = 0$, $\Sigma = \textit{Var}^*(y_t | \{y_\tau\}_{-\infty<\tau\leq t-1})$
- $u_t \sim WN(0, \Sigma)$
- $\Psi_0 = I$ *and* $\sum_{k=0}^{\infty} \Psi_k^2 < \infty$ *(*$\Psi(L)$ *is square-summable)*
- d_t *is deterministic and orthogonal to* u_t

The sequences $\{\Psi_k\}$, $\{u_t\}$ *and* $\{d_t\}$ *are unique*

Wold decomposition: discussion

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

  VMA(∞) + deterministic component

- Interpretation of the u's
  - The Wold decomposition splits a process $\{y_t\}$ into one-step-ahead prediction errors and a perfectly predictable residual
  - The u's are also called Wold innovations
  - Note: we can also turn the Wold decomposition into an 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 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:

$$\Psi_l = \text{Cov}(y_t, u_{t-l})\Sigma^{-1}$$

    - It thus follows that the Wold decomposition is **identifiable** from aggregate time series data (just like autocovariances & spectral densities)
    - The Wold decomposition is our third fundamental representation. Second-order properties can be represented by ACF, spectrum and Wold decomposition

Outline

Why is Slutzky-Frisch a good paradigm?

We don't know the DGP of the economy. But we have seen that

1. Most macro models admit a **structural VMA(∞) representation**: a model is nothing but a mapping from shocks ε_t to observables y_t
2. From the **Wold decomposition**, we know that we can write any time series as a VMA(∞) plus some deterministic component
   - We don't know the DGP of the economy, but we know that it has a VMA(∞) representation
   - That representation, however, does not have to be meaningful: only if the Wold innovations u_t and ε_t span the same space, the Wold and the structural VMA coincide

This makes the **impulse-propagation paradigm** a powerful framework to think about economic fluctuations

Summary

- We saw some basic time series 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 additional **economic assumptions** are needed to learn about the Θ's (and thus all our objects of interest)