Lecture 5: ARMA Models and the Wold's Decomposition

Raul Riva

FGV EPGE

October, 2025

Intro

Intro

- So far, we learned what MA and AR models are;
- Very different ways of modelling dependence over time;
- An ARMA model combines both ways of modelling dependence $\implies$ very flexible model;
- We will talk about forecasting with ARMA models as well – really useful in the real world!
- Wold's Decomposition: ARMA models are *the* class of stationary processes you should worry about!

Questions?

ARMA(p, q) Models

ARMA(p,q) Models

- Let ε_t be a white noise with variance σ^2;
- An ARMA(p,q) process y_t satisfies the following dynamics:

$$y_t = \mu + \phi_1 y_{t-1} + ... + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + ... + \theta_q \varepsilon_{t-q}$$

- Equivalently:

$$\phi(L) y_t = \mu + \theta(L) \varepsilon_t$$

  where $\phi(L) = 1 - \phi_1 L - ... - \phi_p L^p$ and $\theta(L) = 1 + \theta_1 L + ... + \theta_q L^q$;

ARMA(p, q) Models

- Let ε_t be a white noise with variance σ^2;
- An ARMA(p, q) process y_t satisfies the following dynamics:

$$y_t = \mu + \phi_1 y_{t-1} + ... + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + ... + \theta_q \varepsilon_{t-q}$$

- Equivalently:

$$\phi(L) y_t = \mu + \theta(L) \varepsilon_t$$

  where $\phi(L) = 1 - \phi_1 L - ... - \phi_p L^p$ and $\theta(L) = 1 + \theta_1 L + ... + \theta_q L^q$;
- What conditions will ensure that y_t is stationary?

When is an ARMA(p, q) stationary?

- If the roots of $\phi(L)$ are outside of the unit circle, then $\phi^{-1}(L)$ is well-defined;
- We can invert $\phi(L)$ to get:

$$y_t = \frac{\mu}{1 - \phi_1 - ... - \phi_p} + \theta(L)\varepsilon_t = \frac{\mu}{1 - \phi_1 L - ... - \phi_p L^p} + \psi(L)\varepsilon_t$$

where $\psi(L) = \theta(L)\phi^{-1}(L)$;

- The stationarity of y_t depends only on the roots of $\phi(L)$, not on the roots of $\theta(L)$;

When is an ARMA(p, q) stationary?

- If the roots of $\phi(L)$ are outside of the unit circle, then $\phi^{-1}(L)$ is well-defined;
- We can invert $\phi(L)$ to get:

$$y_t = \frac{\mu}{1 - \phi_1 - ... - \phi_p} + \theta(L)\varepsilon_t = \frac{\mu}{1 - \phi_1 L - ... - \phi_p L^p} + \psi(L)\varepsilon_t$$

where $\psi(L) = \theta(L)\phi^{-1}(L)$;

- The stationarity of y_t depends only on the roots of $\phi(L)$, not on the roots of $\theta(L)$;
- From here, it is clear that: $\mathbb{E}[y_t] = \frac{\mu}{1-\phi_1-...-\phi_p}$

What kind of autocovariance structure do we get?

- Notice that if $j > q$, then:

$$\gamma_j = \phi_1 \gamma_{j-1} + ... + \phi_p \gamma_{j-p}$$

- Roots of $\phi(L)$ are outside the unit circle $\implies \gamma_j \to 0$ as $j \to \infty$ (exponentially fast!);

What kind of autocovariance structure do we get?

- Notice that if $j > q$, then:

$$\gamma_j = \phi_1 \gamma_{j-1} + ... + \phi_p \gamma_{j-p}$$

- Roots of $\phi(L)$ are outside the unit circle $\implies \gamma_j \to 0$ as $j \to \infty$ (exponentially fast!);
- For $j \leq q$, things are way more complicated (see Section 3.3 in Brockwell and Davis);
- The reason is that the q last shocks can impose very complex dynamics without affecting stationarity!

What kind of autocovariance structure do we get?

- Notice that if $j > q$, then:

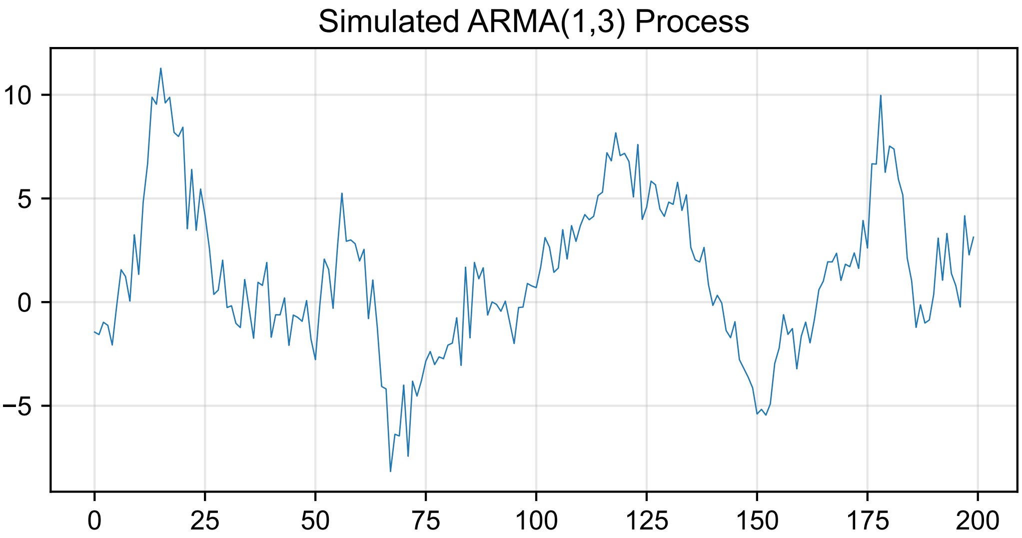
$$\gamma_j = \phi_1 \gamma_{j-1} + \ldots + \phi_p \gamma_{j-p}$$

- Roots of $\phi(L)$ are outside the unit circle $\implies \gamma_j \to 0$ as $j \to \infty$ (exponentially fast!);
- For $j \leq q$, things are way more complicated (see Section 3.3 in Brockwell and Davis);
- The reason is that the q last shocks can impose very complex dynamics without affecting stationarity!

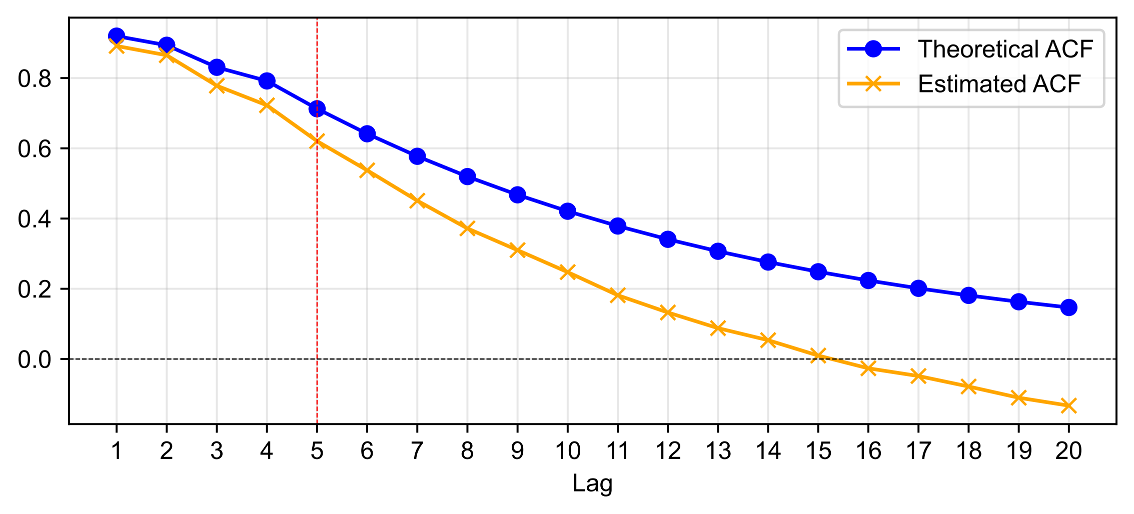
What is important here:

After lag q the decay should be fast and exponential. The MA part will never create trouble for stationarity.

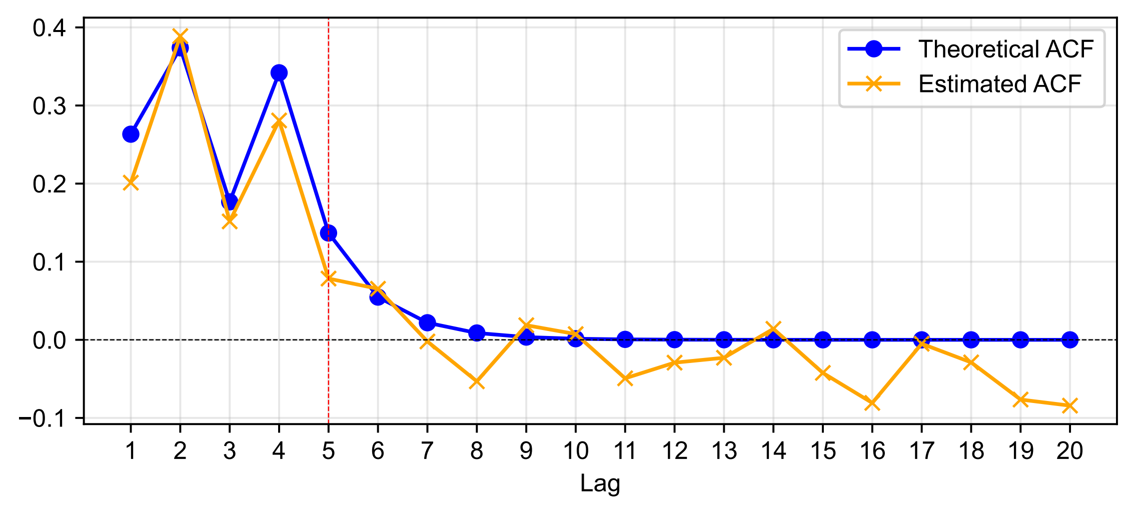
Simulated Example: $y_t = 0.9y_{t-1} + \varepsilon_t - 0.8\varepsilon_{t-1} + 0.6\varepsilon_{t-2} + 0.4\varepsilon_{t-3} + 0.8\varepsilon_{t-4}$



Theoretical vs Estimated ACF



Let's repeat it with $\phi_1 = 0.4$



Example for ARMA(1,1)

- Consider an ARMA(1,1) process:

$$y_t = \mu + \phi_1 y_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1}, \qquad |\phi_1| < 1$$

- We already know that $\gamma_2 = \phi_1 \gamma_1$;
- More generally, $\gamma_j = \phi_1 \gamma_{j-1}$ for $j \geq 2$. We just need to find γ_0 and γ_1;

Example for ARMA(1,1)

- Consider an ARMA(1,1) process:

$$y_t = \mu + \phi_1 y_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1}, \qquad |\phi_1| < 1$$

- We already know that $\gamma_2 = \phi_1 \gamma_1$;
- More generally, $\gamma_j = \phi_1 \gamma_{j-1}$ for $j \geq 2$. We just need to find γ_0 and γ_1;

$$\gamma_0 = \phi_1 \gamma_1 + Cov(\varepsilon_t, y_t) + \theta_1 Cov(\varepsilon_{t-1}, y_t) = \phi_1 \gamma_1 + \sigma^2 + \theta_1 \phi_1 \sigma^2 + \theta_1^2 \sigma^2$$
$$\gamma_1 = \phi_1 \gamma_0 + Cov(\varepsilon_t, y_{t-1}) + \theta_1 Cov(\varepsilon_{t-1}, y_{t-1}) = \phi_1 \gamma_0 + \theta_1 \sigma^2$$

Example for ARMA(1,1)

- Consider an ARMA(1,1) process:

$$y_t = \mu + \phi_1 y_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1}, \qquad |\phi_1| < 1$$

- We already know that $\gamma_2 = \phi_1 \gamma_1$;
- More generally, $\gamma_j = \phi_1 \gamma_{j-1}$ for $j \geq 2$. We just need to find γ_0 and γ_1;

$$\gamma_0 = \phi_1 \gamma_1 + Cov(\varepsilon_t, y_t) + \theta_1 Cov(\varepsilon_{t-1}, y_t) = \phi_1 \gamma_1 + \sigma^2 + \theta_1 \phi_1 \sigma^2 + \theta_1^2 \sigma^2$$
$$\gamma_1 = \phi_1 \gamma_0 + Cov(\varepsilon_t, y_{t-1}) + \theta_1 Cov(\varepsilon_{t-1}, y_{t-1}) = \phi_1 \gamma_0 + \theta_1 \sigma^2$$

Now we can solve this linear system!

Example of ARMA(1,1)

- Just solve the linear system:

$$
\begin{aligned}
\gamma_0 &= \frac{(1+\theta_1^2+2\theta_1\phi_1)\sigma^2}{1-\phi_1^2} \\
\gamma_1 &= \left[\theta+\phi+\frac{(\theta+\phi)^2\phi}{1-\phi^2}\right]\sigma^2 \\
\gamma_j &= \phi_1^{j-1}\gamma_1, \quad j \geq 2
\end{aligned}
$$

Example of ARMA(1,1)

- Just solve the linear system:

$$
\begin{aligned}
\gamma_0 &= \frac{(1+\theta_1^2+2\theta_1\phi_1)\sigma^2}{1-\phi_1^2} \\
\gamma_1 &= \left[\theta+\phi+\frac{(\theta+\phi)^2\phi}{1-\phi^2}\right]\sigma^2 \\
\gamma_j &= \phi_1^{j-1}\gamma_1, \quad j \geq 2
\end{aligned}
$$

- Wait a minute... what would happen if $\theta_1 = -\phi_1$?

The Minimal Representation

- In that case y_t would be white noise!!!

The Minimal Representation

- In that case y_t would be white noise!!!
- In fact, ARMA models can be *overparameterized*;
- Remember that we can always factor the lag polynomials:

$$\phi(L) = (1 - \phi_1 L)(1 - \phi_2 L) \dots (1 - \phi_p L)$$

$$\theta(L) = (1 + \theta_1 L)(1 + \theta_2 L) \dots (1 + \theta_q L)$$

The Minimal Representation

- In that case y_t would be white noise!!!
- In fact, ARMA models can be *overparameterized*;
- Remember that we can always factor the lag polynomials:

$$\phi(L) = (1 - \phi_1 L)(1 - \phi_2 L) \dots (1 - \phi_p L)$$

$$\theta(L) = (1 + \theta_1 L)(1 + \theta_2 L) \dots (1 + \theta_q L)$$

- If there are common roots that cancel out, we can have the exact same correlation structure with $p-1$ lags of y_t and $q-1$ lags of ε_t;
- This is a theoretical justification to prefer small values of p and q when estimating an ARMA model!
- When we write "ARMA(p,q)" we implicitly mean the *minimal representation*!

Questions?

Forecasting with ARMA Models

Forecasting with ARMA Models

- ARMA models are extremely useful for forecasting;
- They are a simple way to capture complicated dynamics;

Forecasting with ARMA Models

- ARMA models are extremely useful for forecasting;
- They are a simple way to capture complicated dynamics;
- Our job: compute conditional expectations of some process y_t;
- Our data: $\{y_1, y_2, \dots, y_{T-1}, y_T\} \implies$ a *strip* of *one* realized path with T observations;

Forecasting with ARMA Models

- ARMA models are extremely useful for forecasting;
- They are a simple way to capture complicated dynamics;
- Our job: compute conditional expectations of some process y_t;
- Our data: $\{y_1, y_2, \ldots, y_{T-1}, y_T\} \implies$ a *strip* of *one* realized path with T observations;
- Exact forecasts will depend on the *infinite* past of the process;
- We will use the *available* information to compute the approximate forecasts;

Forecasting with Infinite Data

- Let's say you have a stationary ARMA model and has already computed the MA(∞) representation:

$$y_t - \mu = \psi(L)\varepsilon_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i} = \epsilon_t + \sum_{i=1}^{\infty} \psi_i \epsilon_{t-i}$$

- As usual: $\psi(L) = \sum_{i=0}^{\infty} \psi_i L^i$, with $\psi_0 = 1$ and $\sum_{i=0}^{\infty} |\psi_i| < \infty$;

Forecasting with Infinite Data

- Let's say you have a stationary ARMA model and has already computed the MA(∞) representation:

$$y_t - \mu = \psi(L)\varepsilon_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i} = \epsilon_t + \sum_{i=1}^{\infty} \psi_i \epsilon_{t-i}$$

- As usual: $\psi(L) = \sum_{i=0}^{\infty} \psi_i L^i$, with $\psi_0 = 1$ and $\sum_{i=0}^{\infty} |\psi_i| < \infty$;
- Let's say you want to compute $\mathbb{E}[y_{t+s}|\varepsilon_t, \varepsilon_{t-1}, ...]$;
- Just using the definition:

$$y_{t+s} - \mu = \varepsilon_{t+s} + \psi_1 \varepsilon_{t+s-1} + ... + \psi_{s-1}\varepsilon_{t+1} + \psi_s \varepsilon_t + \sum_{i=s+1}^{\infty} \psi_i \varepsilon_{t+s-i}$$

Forecasting with Infinite Data

- Let's say you have a stationary ARMA model and has already computed the MA(∞) representation:

$$y_t - \mu = \psi(L)\varepsilon_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i} = \epsilon_t + \sum_{i=1}^{\infty} \psi_i \epsilon_{t-i}$$

- As usual: $\psi(L) = \sum_{i=0}^{\infty} \psi_i L^i$, with $\psi_0 = 1$ and $\sum_{i=0}^{\infty} |\psi_i| < \infty$;
- Let's say you want to compute $\mathbb{E}[y_{t+s}|\varepsilon_t, \varepsilon_{t-1}, ...]$;
- Just using the definition:

$$y_{t+s} - \mu = \varepsilon_{t+s} + \psi_1 \varepsilon_{t+s-1} + ... + \psi_{s-1}\varepsilon_{t+1} + \psi_s \varepsilon_t + \sum_{i=s+1}^{\infty} \psi_i \varepsilon_{t+s-i}$$

- This implies: $\mathbb{E}[y_{t+s}|\varepsilon_t, \varepsilon_{t-1}, ...] = \mu + \sum_{i=s}^{\infty} \psi_i \varepsilon_{t+s-i}$

Forecasting with Infinite Data

- There is a shorthand notation that is useful here. Notice that:

$$\frac{\psi(L)}{L^s} = L^{-s} + \psi_1 L^{1-s)} + \psi_2 L^{2-s} + ... + \psi_{s-1} L^{-1} + \psi_s + \psi_{s+1} L + \psi_{s+2} L^2 ...$$

- The *annihilation operator* denoted by $[.]_+$ only considers the positive powers of L:

$$\left[\frac{\psi(L)}{L^s}\right]_+ \equiv \psi_s + \psi_{s+1} L + \psi_{s+2} L^2 + ...$$

- Now we can write:

$$\mathbb{E}[y_{t+s} | \varepsilon_t, \varepsilon_{t-1}, ...] = \mu + \left[\frac{\psi(L)}{L^s}\right]_+ \varepsilon_t$$

Wait, this feels like cheating...

- We never really observe the shocks ε_t. How can this be useful?

Wait, this feels like cheating...

- We never really observe the shocks ε_t. How can this be useful?
- Sometimes we can actually back them out from infinite data!
- Let's go back to the ARMA(p, q) representation: $\phi(L)(y_t - \mu) = \theta(L)\varepsilon_t$
- $\phi(L) = 1 - \phi_1 L - ... - \phi_p L^p$ and $\theta(L) = 1 + \theta_1 L + ... + \theta_q L^q$;
- If all roots of $\phi(L)$ are outside the unit circle, then $\phi^{-1}(L)$ is well-defined;
- If all roots of $\theta(L)$ are outside the unit circle, then $\theta^{-1}(L)$ is well-defined as well!

Wait, this feels like cheating...

- We never really observe the shocks ε_t. How can this be useful?
- Sometimes we can actually back them out from infinite data!
- Let's go back to the ARMA(p, q) representation: $\phi(L)(y_t - \mu) = \theta(L)\varepsilon_t$
- $\phi(L) = 1 - \phi_1 L - \ldots - \phi_p L^p$ and $\theta(L) = 1 + \theta_1 L + \ldots + \theta_q L^q$;
- If all roots of $\phi(L)$ are outside the unit circle, then $\phi^{-1}(L)$ is well-defined;
- If all roots of $\theta(L)$ are outside the unit circle, then $\theta^{-1}(L)$ is well-defined as well!
- In this case we can write:

$$\theta(L)^{-1}\phi(L)(y_t - \mu) = \varepsilon_t \implies \varepsilon_t = \eta(L)(y_t - \mu)$$

where $\eta(L) = \theta^{-1}(L)\phi(L) = \eta_0 + \eta_1 L + \eta_2 L^2 + \ldots$

The General Formula

- So, for an ARMA(p, q) process with all roots outside the unit circle:

$$\mathbb{E}[y_{t+s}|y_t, y_{t-1}, ...] = \mu + \left[\frac{\psi(L)}{L^s}\right]_+ \eta(L)(y_t - \mu)$$

- In practice, a computer will do the necessary algebra to get the coefficients;
- But it is really important to understand *why* a computer can do that!
- This is also called the **Wiener-Kolmogorov** prediction formula;
- The crucial assumptions are stationarity and invertibility, i.e., roots of $\phi(L)$ and $\theta(L)$ must be outside the unit circle!

Example: AR(1)

- Assume that $(1-\phi L)y_t = \varepsilon_t$ with $|\phi| < 1$;
- Then we know that $\psi(L) = \phi(L)^{-1} = 1 + \phi L + \phi^2 L^2 + \ldots$ and $\eta(L) = \phi(L) = 1 - \phi L$;

Example: AR(1)

- Assume that $(1 - \phi L)y_t = \varepsilon_t$ with $|\phi| < 1$;
- Then we know that $\psi(L) = \phi(L)^{-1} = 1 + \phi L + \phi^2 L^2 + ...$ and $\eta(L) = \phi(L) = 1 - \phi L$;
- The annihilation operator gives:

$$\left[\frac{\psi(L)}{L^s}\right]_+ = \phi^s + \phi^{s+1} L + \phi^{s+2} L^2 + ... = \phi^s (1 + \phi L + \phi^2 L^2 + ...) = \phi^s \left(\frac{1}{1 - \phi L}\right)$$

Example: AR(1)

- Assume that $(1-\phi L)y_t = \varepsilon_t$ with $|\phi| < 1$;
- Then we know that $\psi(L) = \phi(L)^{-1} = 1 + \phi L + \phi^2 L^2 + ...$ and $\eta(L) = \phi(L) = 1 - \phi L$;
- The annihilation operator gives:

$$\left[\frac{\psi(L)}{L^s}\right]_+ = \phi^s + \phi^{s+1}L + \phi^{s+2}L^2 + ... = \phi^s(1 + \phi L + \phi^2 L^2 + ...) = \phi^s\left(\frac{1}{1-\phi L}\right)$$

- So we get:

$$\mathbb{E}[y_{t+s}|y_t, y_{t-1}, ...] = \mu + \phi^s(y_t - \mu)$$

Example: AR(1)

- Assume that $(1 - \phi L)y_t = \varepsilon_t$ with $|\phi| < 1$;
- Then we know that $\psi(L) = \phi(L)^{-1} = 1 + \phi L + \phi^2 L^2 + ...$ and $\eta(L) = \phi(L) = 1 - \phi L$;
- The annihilation operator gives:

$$\left[\frac{\psi(L)}{L^s}\right]_+ = \phi^s + \phi^{s+1} L + \phi^{s+2} L^2 + ... = \phi^s (1 + \phi L + \phi^2 L^2 + ...) = \phi^s \left(\frac{1}{1 - \phi L}\right)$$

- So we get:

$$\mathbb{E}[y_{t+s} | y_t, y_{t-1}, ...] = \mu + \phi^s (y_t - \mu)$$

- What happens when $s \to \infty$? What's the intuition for this result?
- You will do the AR(p) case in the problem set!

Example: MA(q)

- Now let's assume that $y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + ... + \theta_q \varepsilon_{t-q}$;
- Then $\psi(L) = \theta(L) = 1 + \theta_1 L + ... + \theta_q L^q$, and $\eta(L) = \theta(L)^{-1}$;
- If $s > q$, then the annihilation operator gives 0;
- In that case: $\mathbb{E}[y_{t+s} | y_t, y_{t-1}, ...] = \mu$. What's the intuition for this result?

Example: MA(q)

- Now let's assume that $y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + ... + \theta_q \varepsilon_{t-q}$;
- Then $\psi(L) = \theta(L) = 1 + \theta_1 L + ... + \theta_q L^q$, and $\eta(L) = \theta(L)^{-1}$;
- If $s > q$, then the annihilation operator gives 0;
- In that case: $\mathbb{E}[y_{t+s}|y_t, y_{t-1}, ...] = \mu$. What's the intuition for this result?
- In case $s \leq q$, then:

$$\left[\frac{\psi(L)}{L^s}\right]_+ = \theta_s + \theta_{s+1} L + ... + \theta_q L^{q-s}$$

- The final forecast is:

$$\mathbb{E}[y_{t+s}|y_t, y_{t-1}, ...] = \mu + \frac{(\theta_s + \theta_{s+1} L + ... + \theta_q L^{q-s})}{\theta(L)^{-1}}(y_t - \mu)$$

Questions?

Forecasting with *Finite* Data

Forecasting with *Finite* Data

- Bad news about what we did: we assumed we had infinite data. That is never the case...
- There are two main ways of dealing with this problem:
  1. Create an "approximate" forecast;
  2. Use some method that explictly accounts for "missing data";
- The most common method for (2) is something called the "Kalman Filter";
- We don't have time to cover it, but it's super useful in Macro/Finance/estimation of DSGE models, etc!
- Most statistical software use (2) as the method to construct forecasts;
- But learning (1) is instructive and yields the same results is T is large!

How to approximate the forecasts

- Consider an ARMA(p, q) process:

$$y_t - \mu = \phi_1(y_{t-1} - \mu) + \ldots + \phi_p(y_{t-p} - \mu) + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q}$$

- We have data $\{y_1, y_2, \ldots, y_T\}$ and we want to forecast y_{T+h} for $h > 0$.
- The main issue is that we only observe finite data and never observe ε_t;

How to approximate the forecasts

- Consider an ARMA(p, q) process:

$$y_t - \mu = \phi_1(y_{t-1} - \mu) + ... + \phi_p(y_{t-p} - \mu) + \varepsilon_t + \theta_1\varepsilon_{t-1} + ... + \theta_q\varepsilon_{t-q}$$

- We have data $\{y_1, y_2, ..., y_T\}$ and we want to forecast y_{T+h} for $h > 0$.
- The main issue is that we only observe finite data and never observe ε_t;
- Main trick: use the equatio above to back out a series of "estimated" shocks $\hat{\varepsilon}_t$;
- Assume that $\varepsilon_t = 0$ for $t \leq$ (before the sample);
- Also assume that $y_t = \mathbb{E}[y_t] = \mu$ for $t \leq 0$ (before the sample);

How to approximate the forecasts

Now we can back out the shocks recursively:

$$
\begin{aligned}
\hat{\varepsilon}_1 &= y_1 - \mu \\
\hat{\varepsilon}_2 &= y_2 - \mu - \phi_1(y_1 - \mu) - \theta_1\hat{\varepsilon}_1 \\
\hat{\varepsilon}_3 &= y_3 - \mu - \phi_1(y_2 - \mu) - \phi_2(y_1 - \mu) - \theta_1\hat{\varepsilon}_2 - \theta_2\hat{\varepsilon}_1 \\
&\vdots \\
\hat{\varepsilon}_q &= y_q - \mu - \phi_1(y_{q-1} - \mu) - \ldots - \phi_{q-1}(y_1 - \mu) - \theta_1\hat{\varepsilon}_{q-1} - \ldots - \theta_{q-1}\hat{\varepsilon}_1 \\
&\vdots \\
\hat{\varepsilon}_T &= y_T - \mu - \phi_1(y_{T-1} - \mu) - \ldots - \phi_p(y_{T-p} - \mu) - \theta_1\hat{\varepsilon}_{T-1} - \ldots - \theta_q\hat{\varepsilon}_{T-q}
\end{aligned}
$$

- Essentially, we are doing $\hat{\varepsilon}_t \equiv y_t - \hat{y}_{t|t-1}$, where $y_{t|t-1}$ denotes the conditional expectation of y_t given the past observations *and* given that we set presample data to the unconditional mean.

How to approximate the forecasts

- From here on, you can just iterate forward;
- In general (see equation 4.2.25 from Hamilton's book):

$$\hat{y}_{t+h|t} - \mu = \begin{cases} \phi_1(y_{t+h-1|t} - \mu) + \phi_2(y_{t+h-2|t} - \mu) + ... + \phi_p(y_{t+h-p|t} - \mu) \\ \qquad + \theta_h \hat{\varepsilon}_t + ... + \theta_q \hat{\varepsilon}_{t+s-q} & \text{if } h \leq p \\ \phi_1(y_{t+h-1|t} - \mu) + \phi_2(y_{t+h-2|t} - \mu) + ... + \phi_p(y_{t+h-p|t} - \mu) & \text{if } h > p \end{cases}$$

How to approximate the forecasts

- From here on, you can just iterate forward;
- In general (see equation 4.2.25 from Hamilton's book):

$$\hat{y}_{t+h|t} - \mu = \begin{cases} \phi_1(y_{t+h-1|t} - \mu) + \phi_2(y_{t+h-2|t} - \mu) + ... + \phi_p(y_{t+h-p|t} - \mu) \\ \qquad + \theta_h \hat{\varepsilon}_t + ... + \theta_q \hat{\varepsilon}_{t+s-q} & \text{if } h \leq p \\ \phi_1(y_{t+h-1|t} - \mu) + \phi_2(y_{t+h-2|t} - \mu) + ... + \phi_p(y_{t+h-p|t} - \mu) & \text{if } h > p \end{cases}$$

- In practice, compute $y_{t+1|t}$, then $y_{t+2|t}$, then $y_{t+3|t}$, and so on;

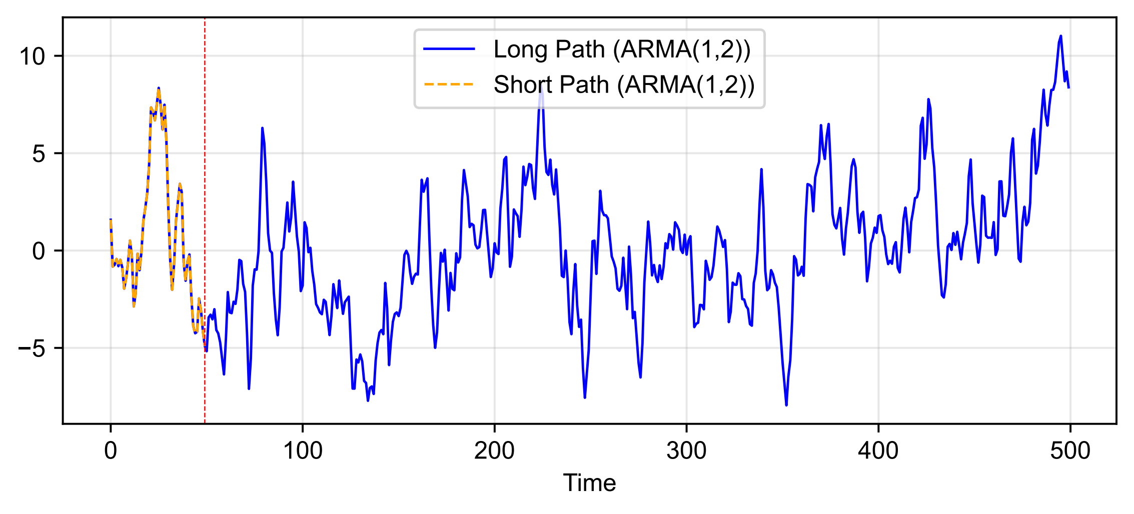
How to approximate the forecasts

- From here on, you can just iterate forward;
- In general (see equation 4.2.25 from Hamilton's book):

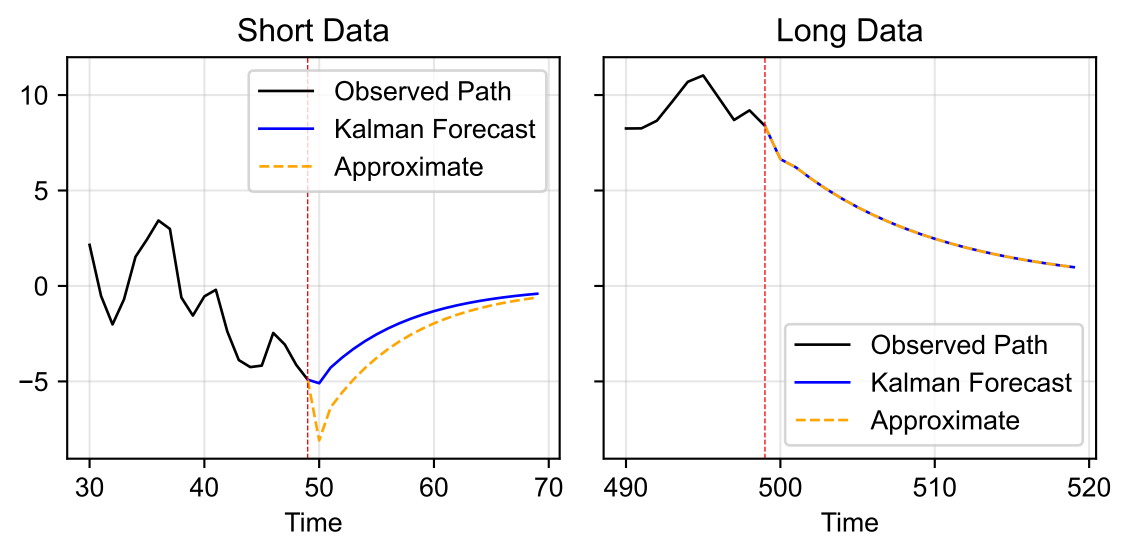
$$\hat{y}_{t+h|t} - \mu = \begin{cases} \phi_1(y_{t+h-1|t} - \mu) + \phi_2(y_{t+h-2|t} - \mu) + ... + \phi_p(y_{t+h-p|t} - \mu) \\ \qquad + \theta_h \hat{\varepsilon}_t + ... + \theta_q \hat{\varepsilon}_{t+s-q} & \text{if } h \leq p \\ \phi_1(y_{t+h-1|t} - \mu) + \phi_2(y_{t+h-2|t} - \mu) + ... + \phi_p(y_{t+h-p|t} - \mu) & \text{if } h > p \end{cases}$$

- In practice, compute $y_{t+1|t}$, then $y_{t+2|t}$, then $y_{t+3|t}$, and so on;
- Notice that, as $h \to \infty$ the forecast approaches μ exponentially fast;
- What's the intuition for this result?
- What is ensuring that the forecast will not explode exponentially fast, by the way?

Example: ARMA(1,2)



Example: ARMA(1,2) - Small vs Large Sample Forecasts



Questions?

The Wold Decomposition

The Wold Decomposition

- Why have we paid so much attention to ARMA models?

The Wold Decomposition

- Why have we paid so much attention to ARMA models?
- The Wold Decomposition Theorem is a powerful result;
- It essentially says that any stationary process has an MA(∞) representation;

The Wold Decomposition

- Why have we paid so much attention to ARMA models?
- The Wold Decomposition Theorem is a powerful result;
- It essentially says that any stationary process has an MA(∞) representation;
- By playing with p and q, we can approximate *any* stationary process **arbitrarily well**;
- In a certain sense, ARMA models are *“dense”* in the space of stationary processes!
- For example: computers approximate real numbers using rationals all the time!

Formal Statement

Theorem (Wold's Decomposition)

Let y_t *be a zero-mean covariance-stationary process. Define* $\mathcal{P}_{t-m}[y_t]$ *as the linear projection of* y_t *into* $\{y_{t-m}, y_{t-m-1}, ...\}$. *Also, let* $e_t = y_t - \mathcal{P}_{t-1}[y_t]$ *be the projection error.*
Then, there exists a unique representation of y_t *as:*

$$Y_t = \mu_t + \sum_{j=0}^{\infty} \psi_j e_{t-j} \tag{1}$$

where e_t *is a white noise process,* $\psi_0 = 1$, $\mu_t = \lim_{m\to\infty} \mathcal{P}_{t-m}[Y_t]$, *and* $\sum_{j=1}^{\infty} \psi_j^2 < \infty$.

- A powerful result: linear representations are *the* way to go for stationary processes;
- ARMA methodology: approximate $\phi(L) = 1 + \phi_1 L + \phi_2 L^2 + ...$ by a *ratio* of two polynomials;

Formal Statement

Theorem (Wold's Decomposition)

Let y_t be a zero-mean covariance-stationary process. Define $\mathcal{P}_{t-m}[y_t]$ as the linear projection of y_t into $\{y_{t-m}, y_{t-m-1}, ...\}$. Also, let $e_t = y_t - \mathcal{P}_{t-1}[y_t]$ be the projection error. Then, there exists a unique representation of y_t as:

$$Y_t = \mu_t + \sum_{j=0}^{\infty} \psi_j e_{t-j} \tag{1}$$

where e_t is a white noise process, $\psi_0 = 1$, $\mu_t = \lim_{m\to\infty} \mathcal{P}_{t-m}[Y_t]$, and $\sum_{j=1}^{\infty} \psi_j^2 < \infty$.

- A powerful result: linear representations are *the* way to go for stationary processes;
- ARMA methodology: approximate $\phi(L) = 1 + \phi_1 L + \phi_2 L^2 + ...$ by a *ratio* of two polynomials;
- *"Any covariance-stationary process is an ARMA process"?* True of False?

The End

References

- Chapter 3 from Hamilton's book for and the basics of ARMA models;
- Chapter 4 from Hamilton's book for Forecasting and Wold's Decomposition;