Financial Time Series

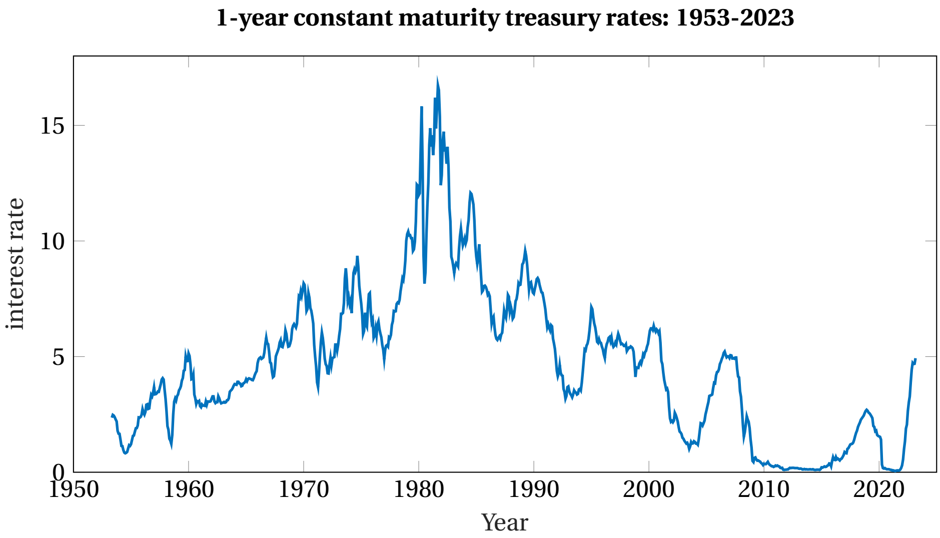
Hui Chen

MIT Sloan

15.450, Spring 2023

Outline

Example: Modeling Interest Rates



Link

Stationarity and Autocorrelation Function

Weak Stationarity

A time series $\{x_t\}$ is weakly stationary if

1. $E(x_t) = \mu$;
2. $E(x_t - \mu)^2 = \gamma_0 < \infty$;
3. $\text{Cov}(x_t, x_{t-j}) = \gamma_j$ for any integer j.

- Weak stationarity is the foundation of statistical inference and forecasting for time series data.

Autocorrelation

For a weakly stationary time series $\{x_t\}$, the correlation between x_t and x_{t-k} is the lag-k autocorrelation of x_t,

$$\rho_k = \frac{\text{Cov}(x_t, x_{t-k})}{\sqrt{Var(x_t)\,Var(x_{t-k})}} = \frac{\gamma_k}{\gamma_0}$$

Autoregressive Models

- AR model: Regression with lagged variables

$$AR(1): \quad x_{t+1} = \phi_0 + \phi_1 x_t + \varepsilon_{t+1}, \quad \varepsilon_{t+1} \sim \mathcal{N}(0, \sigma^2)$$

  ↪ ε_t: Gaussian white noise

- Model of 1-year treasury rates (using data from 1953-2020):

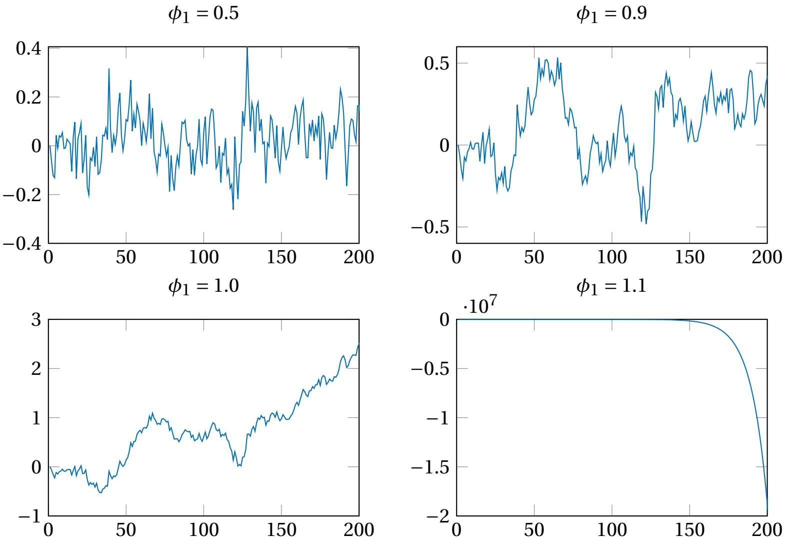
$$r_{t+1} = 0.033 + 0.993\, r_t + \varepsilon_{t+1}$$

Properties of AR(1)

- Some basic properties of AR(1):
  - ↪ Stationarity: $|\phi_1| < 1$
  - ↪ Mean: $E(r_t) = \frac{\phi_0}{1-\phi_1} = \mu$
  - ↪ Variance: $Var(r_t) = \frac{\sigma^2}{1-\phi_1^2}$
  - ↪ Autocorrelations: $\rho_1 = \phi_1, \rho_2 = \phi_1^2, \ldots, \rho_k = \phi_1^k$
- De-meaned representation:

$$r_{t+1} - \mu = \phi_1 (r_t - \mu) + \varepsilon_{t+1}$$

Stationarity of AR(1)



Forecasting with AR(1)

- 1-step ahead forecast:

$$E_t[r_{t+1}] = \phi_0 + \phi_1 r_t = \mu + \phi_1 (r_t - \mu)$$

  ↪ Mean reversion: $|\phi_1| < 1$

- 1-step ahead forecast error:

$$r_{t+1} - E_t[r_{t+1}] = \varepsilon_{t+1}$$

  ↪ Q: Distribution of 1-step ahead forecast error?

- ℓ-step ahead forecast:

$$E_t[r_{t+\ell}] - \mu = \phi_1^\ell (r_t - \mu) \quad \Rightarrow \quad E_t[r_{t+\ell}] = \mu + \phi_1^\ell (r_t - \mu)$$

  ↪ Q: Distribution of ℓ-step ahead forecast error?

Half-life

Time τ needed for the average distance from the mean to shrink by half, $|E_t[r_{t+\tau}] - \mu| = \frac{1}{2}|r_t - \mu|$.

$$\tau = -\frac{\ln(2)}{\ln(|\phi_1|)}, \quad |\phi_1| < 1$$

AR and VAR Models

- AR(p):

$$x_{t+1} = \phi_0 + \phi_1 x_t + \cdots + \phi_p x_{t+1-p} + \varepsilon_{t+1}, \quad \varepsilon_{t+1} \sim \mathcal{N}(0, \sigma^2)$$

- VAR(p) (**V**ector **A**uto**R**egressive model):

$$\mathbf{x}_{t+1} = a_0 + A_1 \mathbf{x}_t + \cdots + A_p \mathbf{x}_{t+1-p} + \varepsilon_{t+1}, \quad \varepsilon_{t+1} \sim \mathcal{N}(0, \Omega)$$

where $\mathbf{x}_t$ and a_0 are N-dim vectors, A_n are $N \times N$ matrices, and ε_t is N-dim vector of shocks.

Example: VAR(1)

- Modeling multiple series jointly to study the relationship between the series, and to improve the accuracy of forecasts.

- Example: real GDP growth (g_t) and inflation (π_t)

$$\begin{bmatrix} g_{t+1} \\ \pi_{t+1} \end{bmatrix} = \begin{bmatrix} a_g \\ a_\pi \end{bmatrix} + \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \begin{bmatrix} g_t \\ \pi_t \end{bmatrix} + \begin{bmatrix} \varepsilon_{g,t+1} \\ \varepsilon_{\pi,t+1} \end{bmatrix}$$

- Example: forecasting excess returns (r_t) using dividend-price ratio (x_t)

$$\begin{bmatrix} r_{t+1} \\ x_{t+1} \end{bmatrix} = \begin{bmatrix} a_r \\ a_x \end{bmatrix} + \begin{bmatrix} 0 & b \\ 0 & \rho \end{bmatrix} \begin{bmatrix} r_t \\ x_t \end{bmatrix} + \begin{bmatrix} \varepsilon_{r,t+1} \\ \varepsilon_{x,t+1} \end{bmatrix}$$

Moving-average Models

- MA model: Shocks with finite life

$$MA(1): \quad x_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

  - ↪ Always stationary
  - ↪ Autocorrelated shocks
  - ↪ Finite memory: $\rho_\ell = 0$ for $\ell > 1$

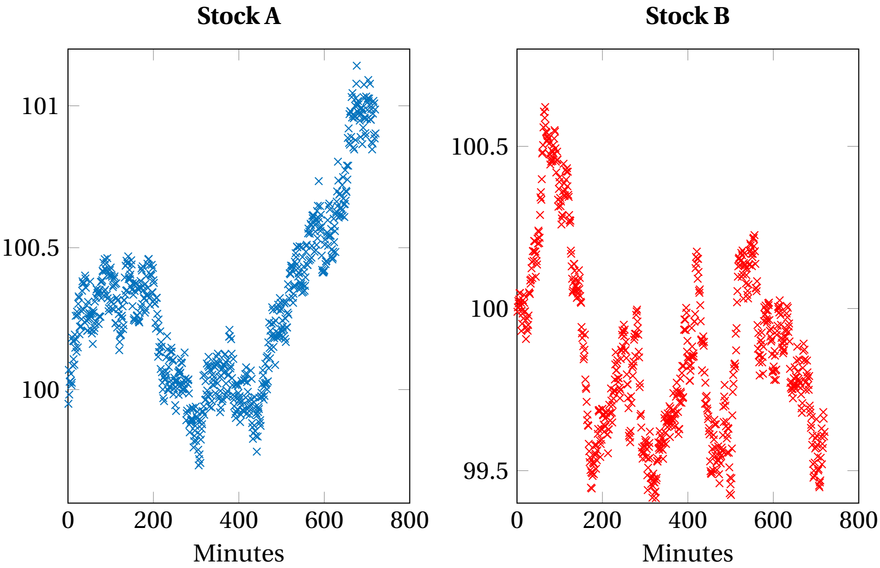
- MA(q):

$$x_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}$$

- Combing AR and MA: ARMA

$$ARMA(1,1): \quad x_{t+1} = \phi_0 + \phi_1 x_t + \varepsilon_{t+1} - \theta \varepsilon_t$$

Which stock is more liquid?



Example of MA(1): Bid-Ask Bounce

- Assume market makers buy at the bid price P_b and sell at the ask price P_a.
- Observed market price P_t is assumed to be

$$P_t = P^* + I_t \frac{S}{2}$$

  - ↪ P^*: fundamental value in a frictionless market (assumed to be constant)
  - ↪ $S = P_a - P_b$: bid-ask spread
  - ↪ I_t: iid Bernoulli (taking value of 1 and -1 with probability 0.5)
  - ↪ Interpretation: buyer-initiated ($I_t = 1$) vs. seller-initiated transaction ($I_t = -1$)
- Observed price change:

$$\Delta P_t = (I_t - I_{t-1}) \frac{S}{2}$$

- Autocovariance:

$$\mathrm{Cov}(\Delta P_t, \Delta P_{t-j}) = \begin{cases} S^2/2 & \text{if } j = 0 \\ -S^2/4 & \text{if } j = 1 \\ 0 & \text{if } j > 1 \end{cases}$$ Why negative?

- Autocovariance of returns as an indicator of illiquidity!

Nonstationary Time Series

- Random walk (with drift):

$$x_{t+1} = a + x_t + \varepsilon_{t+1}$$

  - ↪ A popular model for log stock prices.
  - ↪ Nonstationary (unit root) ⇒ First difference becomes white noise.
  - ↪ Permanent shocks.
  - ↪ Time trend:

$$x_t = at + x_0 + \varepsilon_t + \cdots + \varepsilon_1$$

- Trend-stationary time series:

$$x_t = \beta_0 + \beta_1 t + y_t$$

where y_t is a stationary time series, e.g., an AR(1).

- Unit-root test (Dickey-Fuller test):

$$x_{t+1} = \phi_0 + \phi_1 x_t + \varepsilon_{t+1}$$

  - ↪ $H_0 : \phi_1 = 1$ vs. $H_a : \phi_1 < 1$

$$DF = \frac{\widehat{\phi}_1 - 1}{SE(\widehat{\phi}_1)}$$

Outline

MLE for Dependent Observations

- MLE approach works even if observations are dependent.
- Need dependence to die out quickly enough.
- Consider a time series $x_t, x_{t+1}, \ldots$ and assume that the distribution of x_{t+1} depends only on L lags: $x_t, \ldots, x_{t+1-L}$.
- Likelihood function

$$p(\mathbf{x}|\theta) = p(x_1|\theta)\,p(x_2|x_1;\theta)\,p(x_3|x_2,x_1;\theta)\ldots p(x_L|x_{L-1},\ldots,x_1;\theta)$$
$$\prod_{t=L+1}^{T} p(x_t|x_{t-1},\ldots,x_{t-L};\theta)$$

- θ maximizes the (conditional) likelihood

$$L(x_{L+1},\ldots,x_T|x_L,\ldots,x_1;\theta) \equiv \prod_{t=L+1}^{T} p(x_t|x_{t-1},\ldots,x_{t-L};\theta)$$

  if T is large and x_t is stationary.
- Short-hand notation:

$$L(\theta) \;\text{ and }\; \mathscr{L}(\theta) \equiv \ln L(\theta)$$

MLE for AR(p) Time Series

- AR(p) with IID Gaussian errors:

$$x_{t+1} = a_0 + a_1 x_t + \cdots + a_p x_{t+1-p} + \varepsilon_{t+1}, \quad \varepsilon_{t+1} \sim \mathcal{N}(0, \sigma^2)$$

- Conditional on $(x_t, \ldots, x_{t+1-p})$, x_{t+1} is Gaussian with mean $a_0 + a_1 x_t + \cdots + a_p x_{t+1-p}$ and variance σ^2.

- Construct likelihood:

$$\mathcal{L}(\theta) = \sum_{t=p}^{T-1} - \ln \sqrt{2\pi\sigma^2} - \frac{(x_{t+1} - a_0 - a_1 x_t - \cdots - a_p x_{t+1-p})^2}{2\sigma^2}$$

- MLE estimates of $(a_0, a_1, \ldots, a_p)$ are the same as OLS:

$$\underset{(a_0, a_1, \ldots, a_p)}{\arg\max}\ \mathcal{L}(\theta) = \underset{(a_0, a_1, \ldots, a_p)}{\arg\min} \sum_{t=p}^{T-1} (x_{t+1} - a_0 - a_1 x_t - \cdots - a_p x_{t+1-p})^2$$

MLE for VAR(p) Time Series

- VAR(p) with IID Gaussian errors:

$$x_{t+1} = a_0 + A_1 x_t + \cdots + A_p x_{t+1-p} + \varepsilon_{t+1}, \quad \varepsilon_{t+1} \sim \mathcal{N}(0, \Omega)$$

  where x_t and a_0 are N-dim vectors, A_n are $N \times N$ matrices, and ε_t are N-dim vectors of shocks.

- Conditional on $(x_t, \ldots, x_{t+1-p})$, x_{t+1} is Gaussian with mean $a_0 + A_1 x_t + \cdots + A_p x_{t+1-p}$ and cov matrix Ω.
- Construct conditional log-likelihood:

$$\mathcal{L}(\theta) = \sum_{t=p}^{T-1} -\ln\sqrt{(2\pi)^N |\Omega|} - \frac{1}{2}\varepsilon'_{t+1}\Omega^{-1}\varepsilon_{t+1}$$

MLE for VAR(p) Time Series

- Parameter estimation:

$$\max_{a_0,A_1,\ldots,A_p,\Omega} \mathscr{L}(\theta) \Leftrightarrow \min_{a_0,A_1,\ldots,A_p,\Omega} \sum_{t=p}^{T-1} \ln\sqrt{(2\pi)^N|\Omega|} + \frac{1}{2}\varepsilon'_{t+1}\Omega^{-1}\varepsilon_{t+1}$$

- Optimality conditions for $a_0, A_1, ..., A_p$:

$$\sum_t \left[x_{t-i}\varepsilon'_{t+1}\right] = 0,\ i = 0, 1, \ldots, p-1, \quad \sum_t \varepsilon_{t+1} = 0$$

  where

$$\varepsilon_{t+1} = x_{t+1} - \left(a_0 + A_1 x_t + \cdots + A_p x_{t+1-p}\right)$$

- VAR coefficients can be estimated by OLS, equation by equation.

MLE and Model Selection

- In practice, we often do not know the exact model.
- In some situations, MLE can be adapted to perform model selection.
- Suppose we are considering several alternative models, one of which is the correct model.
- If the sample is large enough, we can identify the correct model by comparing maximized likelihoods and penalizing them for the number of parameters they use.
- Various forms of penalties have been proposed, defining various *information criteria*.

VAR(p) Model Selection

- To build a VAR(p) model, we must decide on the order p.
- Without theoretical guidance, use an information criterion.
- Consider two most popular information criteria:

Akaike (AIC) and Bayesian (BIC)

- Each criterion chooses p to maximize the log likelihood subject to a penalty for model flexibility (free parameters). Various criteria differ in the form of penalty.
- In Python, *auto.arima()* can automatically select the parameters using one of the two criteria:

Code

```
from pmdarima.arima import auto_arima
auto_arima(y, max_p = 5, information_criterion = "aic", trace = TRUE)
```

AIC and BIC

- Start by specifying the maximum possible order $\overline{p}$.
- Make sure that $\overline{p}$ grows with the sample size, but not too fast:

$$\lim_{T\to\infty} \overline{p} = \infty, \quad \lim_{T\to\infty} \frac{\overline{p}}{T} = 0$$

For example, can choose $\overline{p} = \frac{1}{4}(\ln T)^2$. Q: Does this satisfy the above conditions?

- Find the optimal VAR order $p^\star$ as

$$p^\star = \underset{0\le p\le \overline{p}}{\arg\max}\, \frac{2}{T}\mathscr{L}(\theta; p) - \text{penalty}(p)$$

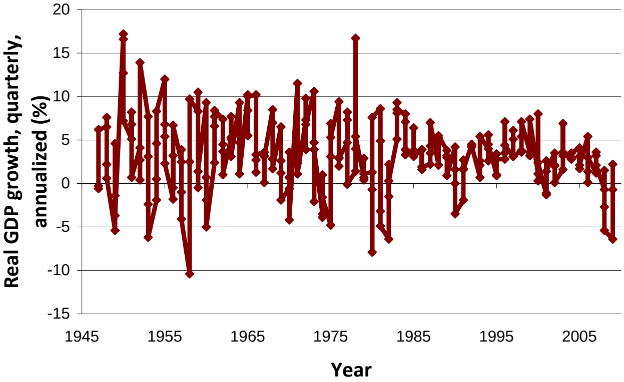
where

$$\text{penalty}(p) = \begin{cases} \text{AIC:} & \frac{2}{T} pN^2 \\ \text{BIC:} & \frac{\ln T}{T} pN^2 \end{cases}$$

- In larger samples, BIC selects lower-order models than AIC.

Example: AR(p) Model of Real GDP Growth

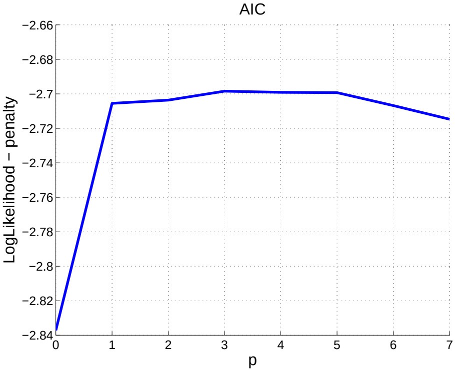
- Model quarterly seasonally adjusted GDP growth (annualized rates).
- Want to select and estimate an AR(p) model.



Source: U.S. Department of Commerce, Bureau of Economic Analysis. National Income and Product Accounts.

Example: AR(p) Model of GDP Growth

- Set $\overline{p} = 7$.

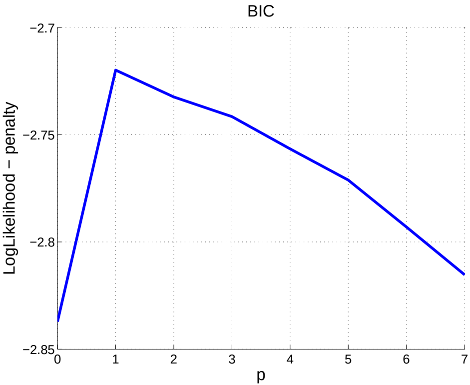


- AIC dictates $p = 3$.
- AR coefficients $a_1, \ldots, a_3$:

$$0.3363,\ 0.1371,\ -0.1165$$

Example: AR(p) Model of GDP Growth

- Set $\overline{p} = 7$.



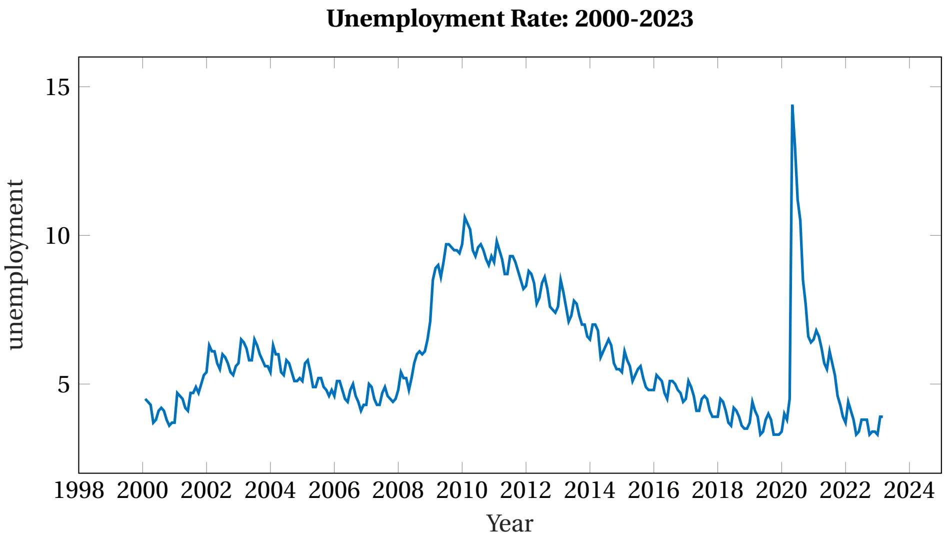
- BIC dictates $p = 1$.
- AR coefficient $a_1 = 0.3611$.

Outline

Seasonal Time Series

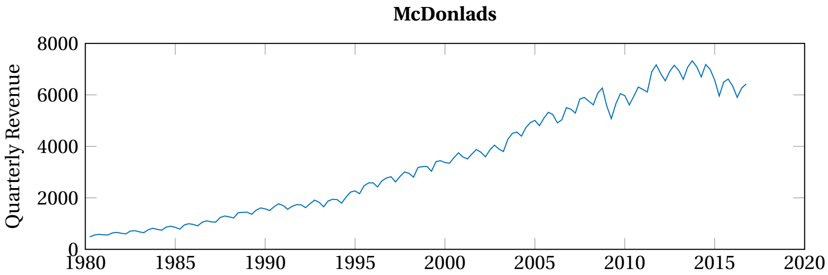
- Some financial time series exhibit seasonal (periodic) patterns.
  - ↪ Examples: unemployment rate, corporate earnings, sales ...
- Cycle length could range from seconds to months and quarters to years.
- Introduces serial correlation/predictability in a particular manner.

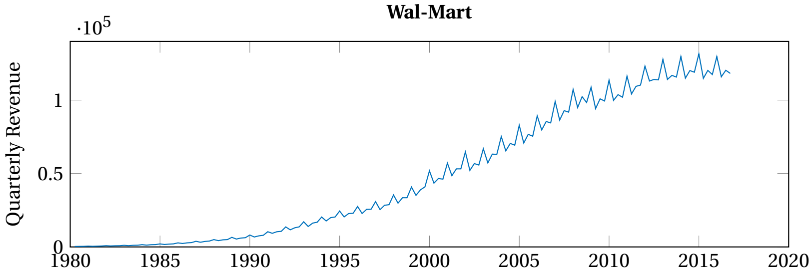
Example: Unemployment



Link

Example: Quarterly revenue





Example: IBM and Coke prices on Jul 19, 2012

Differencing for Seasonal Time Series

- Work with log revenue: $x_t = \log(\text{quarterly revenue})$
- Focus on growth rates, not levels. (Why?)

$$\begin{aligned} \text{regular differencing:} \quad & \Delta x_t = x_t - x_{t-1} \\ \text{seasonal differencing:} \quad & \Delta_4 x_t = x_t - x_{t-4} \end{aligned}$$

- Compare growth rates year over year, not quarter over quarter.

$$\Delta_4(\Delta x_t) = (x_t - x_{t-1}) - (x_{t-4} - x_{t-5})$$

- More generally, for seasonal time series with periodicity s, use Δ_s.

"Airline Model" for Corporate Revenue

Airline model

$$(x_t - x_{t-1}) - (x_{t-s} - x_{t-s-1}) = (\varepsilon_t - \theta_1 \varepsilon_{t-1}) - \theta_s(\varepsilon_{t-s} - \theta_1 \varepsilon_{t-s-1})$$

- s is the periodicity of the series;
- ε_t is white noise.

Q: What are the meanings of θ_1 and θ_s?
Q: What is the ℓ-step ahead forecast $\mathrm{E}_t[x_{t+\ell}]$?

- Multiplicative model: Regular (quarterly) dependence and seasonal (year-to-year) dependence are assumed to be orthogonal.

Modeling McDonalds quarterly revenue

- The goal of this example: Estimate two different time-series models and compare their forecasting performances.

1. Airline model:

$$(x_t - x_{t-1}) - (x_{t-4} - x_{t-5}) = (\varepsilon_t - 0.165\varepsilon_{t-1}) - 0.491(\varepsilon_{t-4} - 0.165\varepsilon_{t-5})$$

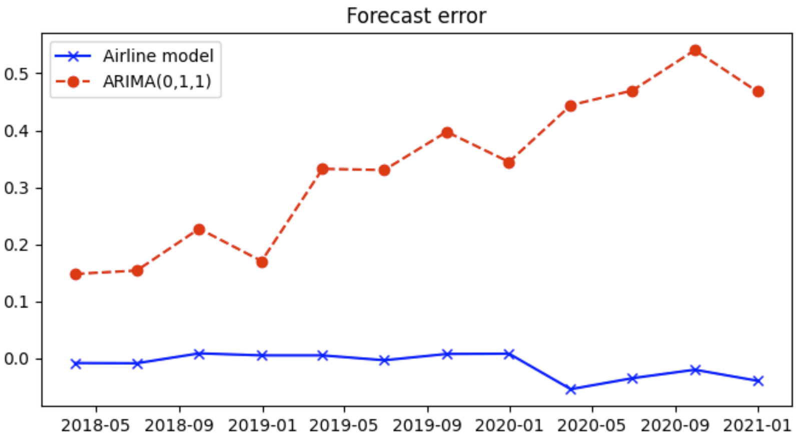
2. A “naive” ARIMA(0,1,1):

$$x_t - x_{t-1} = 0.046 + \varepsilon_t - 0.468\varepsilon_{t-1}$$

Code

```
from statsmodels.tsa.statespace.sarimax import SARIMAX
m1 = SARIMAX(rev, order = (0,1,1), seasonal_order = (0,1,1,4))
m2 = SARIMAX(rev, order = (0,1,1), trend="c")
```

Comparing forecasting performances



Summary

- Weak stationarity
- Basic properties of AR, MA, random walk ...
- Estimating time series models and model selection
- Forecasting time series with seasonality