Time Series Analysis

Models for Non-Stationary Time Series

Hossein Moradi Rekabdarkolaee

South Dakota State University

email: *hossein.moradirekabdarkolaee@sdstate.edu*.

Office: CAME Building, *Room: 240*.

Non-Stationary Time Series

- Any time series without a constant mean over time is non-stationary.
- Models of the form

$$Y_t = \mu_t + \epsilon_t, \qquad E(\epsilon_t) = 0$$

where μ_t is a nonconstant mean function and ϵ_t is a zero-mean, and stationary series.

Exploratory Data Analysis

- Check if a given time series is stationary:
  - μ_t is constant,
  - σ_t^2 is constant,
  - $\gamma_{s,t}$ is function of $k = |t - s|$.
- If not, try to make it stationary using some of the following methods:
  - De-trending,
  - Differencing,
  - Transformations.

De-trending

- Plot time series and/or its moving average: if μ_t seems changing, try to estimate μ_t and subtract it from series

$$\underbrace{Y_t}_{\text{Observed TS}} = \underbrace{\mu_t}_{\text{Trend}} + \underbrace{\epsilon_t}_{\text{Statinary 0-mean TS}}, \qquad \text{estimate trend } \hat{\mu}_t$$

- Then work with

$$\underbrace{\hat{\epsilon}_t}_{\text{de-trended TS}} = \underbrace{Y_t}_{\text{Observed TS}} - \hat{\mu}_t$$

- Since $E[Y_t] = \mu_t$, we have $E[Y_t - \hat{\mu}_t] \approx 0$
- How do we estimate μ_t?
- For deterministic trend (i.e. function of t), we can use regression to estimate μ_t.

Stationarity Through Differencing

- Alternative to fitting linear trend is differencing:

$$Y_t = a + bt + \epsilon_t, \qquad \Rightarrow \quad (Y_t - Y_{t-1} = b + (\epsilon_t - \epsilon_{t-1}))$$

- Note that if $\{\epsilon_t\}$ be stationary then $\{\epsilon_t - \epsilon_{t-1}\}$ is also stationary.
- Differencing is more appropriate when $\{\epsilon_t\}$ is random walk then $\{Y_t\}$ is random walk with drift.
  - Use when $\{Y_t\}$ seems to "hover" around linear trend
  - In this case, de-trending is not enough to make series stationary

Stationarity Through Differencing

- Define first order difference operator ∇ as

$$\nabla Y_t = Y_t - Y_{t-1}$$

- We can extend the differencing to higher (d^{th}) order differences.
- For example when $d = 2$, we have

$$\nabla^2 Y_t = Y_t - 2Y_{t-1} + Y_{t-2}$$

Differencing vs De-trending

- One advantage of differencing over detrending to remove trend is that no parameters are estimated in taking differences.
- One disadvantage of differencing is that it does not provide an "estimate" of the error process $\{\epsilon_t\}$
- If an estimate of the error process is crucial, detrending may be more appropriate.
- If the goal is only to coerce the data to stationarity, differencing may be preferred.

Stationarity Through Differencing

- Consider again the AR(1) model

$$Y_t = \phi Y_{t-1} + \epsilon_t$$

- We have seen that assuming ϵ_t is a true innovation (that is, ϵ_t is uncorrelated with $Y_{t-1}, Y_{t-2}, \ldots$), we must have $|\phi| < 1$.
- How about if $|\phi| > 1$?
- For instance, let $\phi = 3$, then we have

$$Y_t = \sum_{i=0}^{t-1} 3^i \epsilon_{t-i} + 3^t Y_0$$

Stationarity Through Differencing

- Consider a non-stationary AR(1) model with $\phi = 1$ (Random walk)

$$Y_t = Y_{t-1} + \epsilon_t$$

- We can rewrite this as follows and make it a stationary process ($\{\epsilon_t\}$).

$$\nabla Y_t = Y_t - Y_{t-1} = \epsilon_t$$

Stationarity Through Differencing

- Now consider a different approach. Let

$$Y_t = \mu_t + \epsilon_t^*$$

where μ_t is either a deterministic or a stochastic series that changes slowly over time.

- Suppose μ_t is approximately constant over every two consecutive time points.
- Then, we can estimate μ_t at t by choosing β_0 that minimized the following

$$\sum_{j=0}^{1}(Y_{t-j} - \beta_0)^2 \quad \Rightarrow \quad \hat{\mu}_t = \frac{1}{2}(Y_t + Y_{t-1})$$

then **detrended** series at time t is

$$Y_t - \hat{\mu}_t = Y_t - \frac{1}{2}(Y_t + Y_{t-1}) = \frac{1}{2}(Y_t - Y_{t-1}) = \frac{1}{2}\nabla Y_t$$

Stationarity Through Differencing

- A second set of assumptions might be that μ_t is stochastic and changes slowly over time by a random walk model.
- That is

$$Y_t = \mu_t + \epsilon_t \quad \text{with} \quad \mu_t = \mu_{t-1} + \epsilon_t^* \quad \underbrace{\{\epsilon_t\} \;\&\; \{\epsilon_t^*\}}_{\text{independent}} \sim WN$$

then,

$$\nabla Y_t = \nabla \mu_t + \nabla \epsilon_t = \epsilon_t^* + \epsilon_t - \epsilon_{t-1}$$

which has the autocorrelation function of an MA(1) model with

$$\rho_1 = -\frac{1}{2 + \sigma_{\epsilon^*}^2 / \sigma_\epsilon^2}$$

- In either of these situations, we are led to the study of ∇Y_t as a stationary process.

ARIMA Models

- A time series $\{Y_t\}$ is called an **integrated autoregressive moving average (ARIMA)** model if the d^{th} difference $W_t = \nabla^d Y_t$ is a stationary ARMA process.
- If W_t follows an ARMA(p, q) model, then Y_t is an ARIMA(p, d, q).

$$W_t = \nabla^d Y_t \sim \text{ARMA}(p, q)$$

- Y_t is non-stationary, however $\nabla^d Y_t$ is stationary.
- **ARIMA($p, 1, q$) process**.

$$W_t = \sum_{i=1}^{p} \phi_i W_{t-i} + \epsilon_t - \sum_{j=1}^{q} \theta_j \epsilon_{t-j}$$

$$\Rightarrow \quad Y_t - Y_{t-1} = \sum_{i=1}^{p} \phi_i (Y_{t-i} - Y_{t-1-i}) + \epsilon_t - \sum_{j=1}^{q} \theta_j \epsilon_{t-j}$$

ARIMA Models

- This can be rewritten as (difference equation form)

$$Y_t = (1+\phi_1)Y_{t-1} + \sum_{i=2}^{p}(\phi_i - \phi_{i-1})Y_{t-i} - \phi_p Y_{t-p-1} + \epsilon_t - \sum_{j=1}^{q}\theta_j \epsilon_{t-j}$$

- Notice that it appears to be an ARMA($p+1, q$) process, with the following characteristic equation

$$\begin{aligned} 1 - (1+\phi_1)x - \sum_{i=2}^{p}(\phi_i - \phi_{i-1})x^i - \phi_p x^{p+1} \\ = (1 - \phi_1 x - \phi_2 x^2 - \ldots - \phi_p x^p)(1-x) \end{aligned}$$

- this clearly shows one of the roost is $x = 1$, which implies non-stationarity

ARIMA($p, 1, q$) Models

- Explicit representations of the observed series in terms of either W_t or the white noise series underlying W_t are more difficult than in the stationary case.
- Since nonstationary processes are not in statistical equilibrium, we cannot assume that they go infinitely into the past i.e. start at $t = -\infty$.
- However, we can and shall assume that they start at some time point $t = -m$. For convenience, we take $Y_t = 0$ for $t < -m$.
- Let

$$Y_j - Y_{j-1} = W_j$$

- Summing both sides from $j = -m$ to $j = t$, we get

$$Y_t = \sum_{j=-m}^{t} W_j$$

ARIMA($p, 2, q$) Models

- For ARIMA($p, 1, q$) process we obtain

$$Y_t = \sum_{j=-m}^{t} W_j$$

- The ARIMA($p, 2, q$) process can be dealt with similarly by summing twice from the following equation as follows representations

$$Y_j - Y_{j-1} = W_j$$

$$\begin{aligned} Y_t &= \sum_{j=-m}^{t} \sum_{i=-m}^{j} W_i \\ &= \sum_{j=0}^{t+m} (j+1) W_{t-j} \end{aligned}$$

Integrated moving average (IMA(d, q))

- IMA(1, 1): $Y_t = Y_{t-1} + \epsilon_t - \theta\epsilon_{t-1}$
- Then, we have

$$Y_t - Y_{t-1} = W_t = \epsilon_t - \theta\epsilon_{t-1}$$

- After using $Y_t = \sum_{j=-m}^{t} W_j$, and a rearrangement, we have

$$Y_t = \epsilon_t + (1-\theta)\epsilon_{t-1} + (1-\theta)\epsilon_{t-2} + \ldots + (1-\theta)\epsilon_{-m} - \theta\epsilon_{-m-1}$$

- Notice that in contrast to our stationary ARMA models, the weights on the white noise terms **do not die out** as we go into the past.

$$Var(Y_t) = [1 + \theta^2 + (1-\theta)^2(t+m)]\sigma^2$$

$$Corr(Y_t, Y_{t-k}) = \frac{1 - \theta + \theta^2 + (1-\theta)^2(t+m)}{\sqrt{Var(Y_t)Var(Y_{t-k})}} \approx \sqrt{\frac{t+m-k}{t+m}} \approx 1$$

The IMA(2, 2) Model

- For IMA(2, 2) model, we have

$$\nabla^2 Y_t = \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2}$$
$$\Rightarrow Y_t = 2Y_{t-1} - Y_{t-2} + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2}$$

- By using $Y_t = \sum_{j=0}^{t+m} (j+1) W_{t-j}$ for IMA(2, 2), we can express Y_t in terms of ϵ_t, $\epsilon_{t-1}, \ldots$, as follows

$$Y_t = \epsilon_t + \sum_{j=1}^{t+m} \psi_j \epsilon_{t-j} - [(t+m+1)\theta_1 + (t+m)\theta_2]\epsilon_{-m-1}$$
$$- (t+m+1)\theta_2 e_{-m-2}$$

- where $\psi_j = 1 + \theta_2 + (1 - \theta_1 - \theta_2)j$ for $j = 1, 2, 3, \ldots, t+m$.
- Once more we see that the ψ-weights do not die out but form a linear function of j.

The ARI(1, 1) Model

- ARI(1, 1): $\qquad Y_t - Y_{t-1} = \phi(Y_{t-1} - Y_{t-2}) + \epsilon_t$.
- This can be written as follows

$$Y_t = (1+\phi)Y_{t-1} - \phi Y_{t-2} + \epsilon_t, \qquad |\phi| < 1$$

- Notice that it looks like a special AR(2) model.
- It can be shown that the ψ-weights can be obtained by

$$\begin{aligned}(1 - \phi_1 x - \phi_2 x^2 - \ldots - \phi_p x^p)(1 + \psi_1 x + \psi_2 x^2 + \psi_3 x^3 + \ldots)\\ = (1 - \theta_1 x - \theta_2 x^2 - \theta_3 x^3 - \ldots - -\theta_q x^q)\end{aligned}$$

- Solve by matching coefficients, we have

$$\begin{aligned}&\psi_0 = 1, \quad \psi_1 = 1 + \phi\\ &\psi_k = (1+\phi)\psi_{k-1} - \phi\psi_{k-2} \qquad \text{for} \quad k \geq 2\end{aligned}$$

- In this case an explicit solution to the recursion is given as

$$\psi_k = \frac{1 - \phi^{k+1}}{1 - \phi} \quad \text{for} \quad k \geq 1$$

Constant Terms in ARIMA Models

- A nonzero constant mean, μ, in a stationary ARMA model $\{W_t\}$ can be accommodated in either of two ways.

$$\begin{aligned} W_t - \mu =& \phi_1(W_{t-1} - \mu) + \phi_2(W_{t-2} - \mu) + \ldots + \phi_p(W_{t-p} - \mu) \\ & + \epsilon_t - \theta_1\epsilon_{t-1} - \theta_2\epsilon_{t-2} - \theta_q\epsilon_{t-q} \end{aligned}$$

- Alternatively,

$$\begin{aligned} W_t =& \theta_0 + \phi_1 W_{t-1} + \phi_2 W_{t-2} + \ldots + \phi_p W_{t-p} \\ & + \epsilon_t - \theta_1\epsilon_{t-1} - \theta_2\epsilon_{t-2} - \theta_q\epsilon_{t-q} \end{aligned}$$

- Taking expected values on both sides of this,

$$\mu = \theta_0 + (\phi_1 + \phi_2 + \ldots + \phi_p)\mu \quad \Rightarrow \quad \mu = \frac{\theta_0}{1 - \phi_1 - \phi_2 - \ldots - \phi_p}$$

Nonzero mean on the undifferenced series Y_t

- Consider the IMA$(1, 1)$ case with a constant term:

$$Y_t = Y_{t-1} + \theta_0 + \epsilon_t - \theta\epsilon_{t-1} \quad \text{or} \quad W_t = \theta_0 + \epsilon_t - \theta\epsilon_{t-1}$$

- After using $Y_t = \sum_{j=-m}^{t} W_j$, we have

$$Y_t = \epsilon_t + (1-\theta)\epsilon_{t-1} + (1-\theta)\epsilon_{t-2} + \ldots + (1-\theta)\epsilon_{-m} - \theta\epsilon_{-m-1} + (t + m + 1)\theta_0$$

- Comparing this with IMA$(1, 1)$ without a constant term, we see that we have an added **linear deterministic time trend** $(t + m + 1)\theta_0$ with slope θ_0.
- An equivalent representation of the process would then be

$$Y_t = Y'_t + \beta_0 + \beta_1 t$$

where Y'_t is an IMA$(1, 1)$ series with $E(\nabla Y'_t) = 0$ and $E(\nabla Y_t) = \beta_1$.

Over-differencing

- Over-differencing can introduce artificial dependencies (unnecessary levels of dependency); e. g. , consider

$$Y_t = Y_{t-1} + \epsilon_t$$

then

$$\nabla Y_t = \epsilon_t$$

but

$$\nabla^2 Y_t = \epsilon_t - \epsilon_{t-1} \sim \mathrm{MA}(1)$$

- The variance of the over-differenced process will be larger than that of the original process.
- Over-differencing also creates a non-invertible model.
- Noninvertible models also create serious problems when we attempt to estimate their parameters.

Other Transformations

- **Logarithm transformation:** Differencing can be a useful transformation for achieving stationarity.
- The logarithm transformation is also a useful method in certain circumstances.
- Suppose $Y_t > 0$ for all t such that

$$E(Y_t) = \mu_t \qquad \sqrt{Var(Y_t)} = \mu_t \sigma$$

- Then, using log transformation, and utilizing the Taylor expansion of it around μ_t

$$\log(Y_t) \approx \log(\mu_t) + \frac{Y_t - \mu_t}{\mu_t}$$

$$\Rightarrow \qquad E[\log(Y_t)] = \log(\mu_t), \qquad Var(Y_t) = \sigma^2$$

- If the standard deviation of Y_t is proportional to the level of Y_t, then $\log(Y_t)$ will produce a series with approximately constant variance over time.

Other Transformations

- **Log-Difference transformation:**
- For financial data, diff-log transform has nice interpretation as percentage change.
- Assume

$$Y_t = (1 + p_t)Y_{t-1} \quad \Rightarrow \quad p_t = \frac{Y_t - Y_{t-1}}{Y_{t-1}}$$

- where $100p_t$ is the percentage change (possibly negative) from Y_{t-1} to Y_t. Then

$$\nabla[\log(Y_t)] = \log(Y_t) - \log(Y_{t-1}) = \log(\frac{Y_t}{Y_{t-1}}) = \log(1 + p_t)$$

- For relatively small p_t, to a good approximation, $\log(1 + p_t) \approx p_t$. That is

$$p_t \approx \nabla[\log(Y_t)]$$

- Therefore, $\nabla[\log(Y_t)]$ will be relatively stable and perhaps well-modeled by a stationary process.

Power Transformations

- If time series has non-constant variance, a nonlinear transformation can sometimes help
- For positive series in particular, the Box-Cox family of power transforms can be useful:

$$g(x) = \begin{cases} \frac{x^\lambda - 1}{\lambda}, & \text{for } \lambda \neq 0; \\ \log(x), & \text{for } \lambda = 0; \end{cases}$$

- Try different values of λ, and check which one seems to give best results.

Power Transformations

- The power transformation applies only to positive data values.
- If some of the values are negative or zero, a positive constant may be added to all of the values to make them all positive before doing the power transformation.
- Software allows us to consider a range of lambda values and calculate a log-likelihood value for each lambda value based on a normal likelihood function.
- Software provides the 95% confidence interval for lambda.
- Note that nonlinear transforms can also change the mean function μ_t.