Time Series Analysis in a nutshell

Wentao Zhu

December 31, 2025

Now we have a set of time series data $\{x_t\}, t = 1, ..., n$. We model them as observations of a set of random variables $\{X_t\}, t = 1, ..., n$. For these n random variables, we have only one observation for each one. It is impossible to know the joint distribution of $\{X_1, ..., X_n\}$, and we only care about the first and second order moments of these random variables: $\mu_X(t)$ and $\gamma_X(s,t)$. Our goal is to model these data and thus be able to predict future values $\{x_t\}, t = n+1, ...$ based on past and current values.

We want our series to be stationary so that we can model it as an ARMA(p, q) based on the following theorem.

Theorem 0.1 (Wold Decomposition Theorem) *Any stationary time series can be represented as MA(∞), which can be approximated by an ARMA(p, q) model with appropriate p and q.*

So now the problem becomes how we can transform it to be stationary if the series we have is initially not stationary.

If we assume that the seasonality (if it exists) is deterministic here for simplicity, then there are three possible sources of unstationarity in total:

(1) $I(d)$ (stochastic trend) (2) trend (3) seasonality

[Case A] If there is no unit root, we can decompose each random variable X_t based on our classical decomposition model:

$$X_t = m_t + s_t + Y_t, \quad \forall t$$

where:

- m_t is a trend component. s_t is a seasonal component. Y_t is the residual/noise and $E[Y_t] = 0$.
- $m_t + s_t$ is the deterministic part, and Y_t is the stochastic part.

Eventually we can model X_t like this:

$$X_t = \beta_0 + \beta_1 t + \beta_2 t^2 + ... + s(t) + \phi_1 Y_{t-1} + ... + \phi_p Y_{t-p} + Z_t + \theta_1 Z_{t-1} + ... + \theta_q Z_{t-q}$$

A more common and accurate notation (to avoid a problem of circular definition) should be:

$$Y_t = \phi_1 Y_{t-1} + ... + \phi_p Y_{t-p} + Z_t + \theta_1 Z_{t-1} + ... + \theta_q Z_{t-q}$$

where:

- $Y_t = X_t - [\beta_0 + \beta_1 t + \beta_2 t^2 + ... + s(t)]$
- $\{Z_t\} \sim \text{WN}(0, \sigma^2)$ (uncorrelated)

Note that the centralization already happens in $m_t + s_t$ (β_0 already makes sure that $E[Y_t] = 0$), so we don't include constant term in ARMA model.

When we are in period t, if we input $\hat{z}_{t+1} = 0$, which is the best predictor of the future noise, we can get the output $\hat{x}_{t+1|t}$, which is our best predictor for a future value.

[Case B] If there is a unit root, then the trend is stochastic, and we can no longer use the classical decomposition model to get a stationary series. Instead, we conduct differencing.

Step 1: Visualization and Descriptive Statistics

When we want to analyze a set of time series data, the first step is always plotting the data on a graph to visualize them to get a first understanding of its possible trend, seasonality and whether there might be a heteroscedasticity problem after removing trend and seasonality. If there are noticeable outliers, we should analyze whether they can be removed or not. It's also helpful to compute some descriptive statistics.

Step 2: Unit Root Diagnostics

Before we perform de-trending, we need to know the nature of trend. If there is no unit root, which means the trend is deterministic, then we go to Step 3A. If there is a unit root, which means the trend is stochastic, resulted from integrating shocks, then we go to Step 3B.

Step 3A: Estimate Trend and Seasonality

We estimate $\hat{m}_t + \hat{s}_t$ and subtract them from x_t in order to get a stationary residual series $\{\hat{Y}_t\}$ with mean zero. The first order moment condition is guaranteed via de-trending and de-seasonalizing. Zero mean is guaranteed via the constant term β_0 while de-trending, so no further centralization is needed. However, there might be a heteroscedasticity problem where the second order moment condition is not satisfied, which indicates the need for some preliminary transformation like Box-Cox transformation before de-trending to control second order moments.

Step 3B: Differencing (ARIMA)

Conduct differencing to remove the stochastic trend. If the stochastic trend is $I(d)$, then we need at least d times differencing to be stationary. A special case is random walk $I(1)$, where we only need to difference once.

The series may contain both $I(d)$ and order-k deterministic trend $m_k(t)$. In this case, we need to difference d times to get $W_t = \nabla^d X_t$. If $d > k$, then $W_t = \eta_t$ which is a stationary process $I(0)$. If $k \geq d$, then $W_t = m_{k-d}(t) + \eta_t$ and we can estimate this power-$(k-d)$ polynomial to de-trend to get the stationary process η_t.

Notice that if this series contains only a deterministic power-k trend, mathematically we can also use differencing ($k+1$ times) to get a stationary process but practically we prefer to de-trend directly (Step 3A).

$I(d)$ can exist with a drift δ. The most common case is a random walk $I(1)$ with a drift. If the sample mean of the differenced series is significantly deviated from zero, then there might be a drift δ, which means centralization (subtracting δ) is needed if we want to use zero-mean ARMA model later. Different from Step 3A where sample mean is guaranteed zero by Least Squares, although the population expectation is zero here, the sample mean may not be zero significantly.

The series may also contain seasonality, which may also be deterministic or stochastic, but for simplicity we don't discuss further about seasonality here. In SARIMA model, we treat seasonality separately.

Step 4: Test $\{Y_t\}$ or $\{W_t\}$

Firstly, we want to make sure that this series $\{Y_t\}$ is stationary.

Secondly, we want to test if there is dependence among them (if they are already white noise) using Ljung-Box test. If no, then these are independent random variables, and we are done (, except to estimate their mean and variance). If yes, then we can utilize this dependence to further model this series as an ARMA(p, q). This is good for better forecasting as dependence means that past observations of the noise sequence can assist in predicting future values.

Step 5: Model $\{Y_t\}$ or $\{W_t\}$ as ARMA(p, q)

Now we have a set of observed values of $\{Y_t\}$: $\{\hat{y}_t\}$, where $\hat{y}_t = x_t - \hat{\beta}_0 - \hat{\beta}_1 t - \hat{\beta}_2 t^2 - ... - \hat{s}_t$ or $\{W_t\}$: $\{\hat{w}_t\}$.

We need to model $\{Y_t\}$ (or $\{W_t\}$) as a zero-mean ARMA(p, q) based on these observations.

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + ... + \theta_q Z_{t-q}$$

where:

- $\{Z_t\} \sim \text{WN}(0, \sigma^2)$ (uncorrelated)

If $\Phi(z)$ has a root of or extremely close to one, then it suggests possible under-differencing. If $\Theta(z)$ has a root of or extremely close to one, then it suggests possible over-differencing.

Task 1: choose p and q (order selection)

Task 2: estimate the coefficients $\{\phi_i, i = 1, ..., p\}, \{\theta_i, i = 1, ..., q\}$

Task 3: estimate the white noise variance σ^2

Step 6: Forecast Future Values $\hat{x}_{t+1}, \hat{x}_{t+2}...$

For deterministic trend models, forecast = trend forecast + seasonal forecast + ARMA forecast.

For ARIMA models, forecast based on differenced series integrated back.