Time Series Analysis

Frank Zhong

August 2024

Contents

1 Introduction

1.1 Background

Time series analysis examines data collected at regular intervals over a defined time span. This method focuses on studying patterns in sequentially ordered observations, rather than analyzing sporadic or arbitrarily gathered data points. By consistently recording measurements at set time intervals, analysts can identify trends, cycles, and other temporal characteristics within the dataset. For more detailed examples and R codes, see Time Series Tutorial.

1.2 Objectives

- To understand how time series works and what factors affect a certain variable(s) at different points in time.
- Time series analysis will provide the consequences and insights of the given dataset's features that change over time.
- Supporting to derive the predicting the future values of the time series variable.

2 Time Series Decomposition

2.1 Time Series Components

In time series analysis, we assume the decomposition to be in three parts: the seasonality component S_t, the trend-cycle component $T_t$and the remainder component R_t. Therefore, for the data y_t, we have:

$$y_t = S_t + T_t + R_t$$

However, such an additive decomposition is the most appropriate if the magnitude of the seasonal fluctuations, or the variation around the trend-cycle, does not vary with the level of the time series. For heavier fluctuation of seasonality or variation, we should apply the multiplicative decomposition ($y_t = S_t \times T_t \times R_t$), or we have to stabilise the fluctuations before decomposing.

2.2 Moving Averages

2.2.1 Moving Average Smoothing

We can use the method of moving averages to find the trend component of a data. We take the neighbouring m values and take its average, forming the so-called m-order moving average. Its formula is as follows:

$$\hat{T_t} = \frac{1}{m} \sum_{j=-k}^{k} y_{t+j}$$

Where the order of the moving average $m = 2k + 1$. That is, the estimate of the trend-cycle at time t is obtained by averaging values of the time series within k periods of t. Observations that are nearby in time are also likely to be close in value. Therefore, the average eliminates some of the randomness in the data, leaving a smooth trend-cycle component. We call this an m-MA, meaning a moving average of order m.

Figure 1: Different Orders of MAs.

2.2.2 Moving Averages of Moving Averages

Usually, the m in moving averages is set to be odd. Since we expect our averages to be symmetry (i.e., for odd values of m, the value at time t will be exactly in the middle).

However, we can still make an even-order moving average to be symmetric by taking the moving average of moving averages. Generally, even moving averages of the moving averages are applied to even moving averages and vise versa. The output is often denoted as $m_1 \times m_2$-MA (e.g. 2×4-MA and 3×3-MA are two commonly used ones), representing taking m_1-order moving average after m_2-order moving averages.

2.2.3 Weighted Moving Averages

Sometimes, we'd like to offer different weights to the values at different times. The weighted m-order moving average can be denoted as:

$$\hat{T}_t = \sum_{j=-k}^{k} a_j y_{t+j}$$

Where $k = \frac{m-1}{2}$ and the weights are $[a_{-k}, \cdots, a_k]$. The weights are all symmetric around the middle value, that is, $a_j = a_{-j}$. All weights add up to 1. The general moving average is a special case when all weights are equal (i.e., $a_{-k} = \cdots = a_k = \frac{1}{m}$).

2.3 Classical Time Series Decomposition

2.3.1 Algorithms

When decomposing the time series, the first step is to estimate the trend. Two different approaches could be used for this [1]:

- The first approach is to estimate the trend with a smoothing procedure such as moving averages. With this approach, an equation is not used to describe trend.
- The second approach is to model the trend with a regression equation.

The second step is to "de-trend" the series. For an additive decomposition, this is done by subtracting the trend estimates from the series. For a multiplicative decomposition, this is done by dividing the series by the trend values.

Next, seasonal factors are estimated using the de-trended series. For monthly data, this entails estimating an effect for each month of the year. For quarterly data, this entails estimating an effect for each quarter. The simplest method for estimating these effects is to average the de-trended values for a specific season. For instance, to get a seasonal effect for January, we average the de-trended values for all Januaries in the series, and so on.

The final step is to determine the random (remainder) component.

- For the additive model: $random = data - trend - seasonal$.
- For the multiplicative model: $random = \frac{data}{trend \times seasonal}$.

The random component could be analysed for such things as the mean squared size (variance), or possibly even for whether the component is actually random or might be modeled with an ARIMA model.

2.3.2 Drawbacks

- The estimate of the trend-cycle is unavailable for the first few and last few observations. For example, if $m = 12$, there is no trend-cycle estimate for the first six or the last six observations. Consequently, there is also no estimate of the remainder component for the same time periods.
- The trend-cycle estimate tends to over-smooth rapid rises and falls in the data.
- Classical decomposition methods assume that the seasonal component repeats from year to year. For many series, this is a reasonable assumption, but for some longer series it is not. For example, electricity demand patterns have changed over time as air conditioning has become more widespread.
- Occasionally, the values of the time series in a small number of periods may be particularly unusual. For example, the monthly air passenger traffic may be affected by an industrial dispute, making the traffic during the dispute different from usual. The classical method is not robust to these kinds of unusual values.

2.4 X11 Decomposition

2.4.1 Algorithms

The X-11 method is based on an iterative principle of estimation of the different components using appropriate moving averages at each step of the algorithm. A simple X11 seasonal adjustment algorithm can be thought of in eight steps. The steps presented below are designed for the *monthly* time series. For quarterly time series, the 2×4-MA instead of the 2×12-MA is used.
For time series $X_t = C_t + S_t + I_t$ (C_t stands for the trend-cycle component, S_t stands for the seasonality component and I_t stands for the random component)[2]:

- Step 1: estimating the trend-cycle component by a 2×12-MA:

$$C_t^{(1)} = M_{2\times 12}\left(X_t\right)$$

  This 2×12-MA used has coefficients of $\frac{1}{24}\{1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 1\}$, which preserves the linear trends, removes the 12-order constant seasonalities and minimises the variance of random component.

- Step 2: estimating the seasonality-random component:

$$\left(S_t + I_t\right)^{(1)} = X_t - C_t^{(1)}$$

- Step 3: estimating the seasonality component by a 3×3-MA over each month:

$$S_t^{(1)} = M_{3\times 3}\left[(S_t + I_t)^{(1)}\right]$$

Then we normalise the seasonality component so that their sum over each 12-month period is roughly 0.

$$\tilde{S_t}^{(1)} = S_t^{(1)} - M_{2\times 12}\left(S_t^{(1)}\right)$$

- Step 4: estimating the seasonally adjusted series:

$$A_t^{(1)} = (C_t + I_t)^{(1)} = X_t - \tilde{S_t}^{(1)}$$

This first estimation of the seasonally adjusted series must, by construction, contain less seasonality. The X-11 method again executes the algorithm presented above, changing the moving averages to take this property into account.

- Step 5: estimating the trend-cycle component by 13-term Henderson moving average:

$$C_t^{(2)} = H_{13}(A_t^{(1)})$$

Henderson moving averages, while they do not have special properties in terms of eliminating seasonality, are good smoother and keep a locally polynomial trend of degree 2[3]. The Henderson's ideal formula for the Henderson moving average (HMA) of k^{th} order is given by [3]:

$$\sum_{i=t-m}^{t+m} \frac{315\left[(z-1)^2 - i^2\right]\left[z^2 - i^2\right]\left[(z+1)^2 - i^2\right]\left[\left(3z^2 - 16\right) - 11i^2\right]}{8z\left(z^2-1\right)\left(4z^2-1\right)\left(4z^2-9\right)\left(4z^2-25\right)} x_i$$

Where: the order of the average $k = 2z - 3 = 2m + 1$. Hence in the X11 decomposition, for $k = 13$, we have $13 = 2z - 3 = 2m + 1$.

- Step 6: estimating the seasonality-random component:

$$(S_t + I_t)^{(2)} = X_t - C_t^{(2)}$$

- Step 7: estimating the seasonality component by a 3×5-MA over each month:

$$S_t^{(2)} = M_{3\times 5}\left[(S_t + I_t)^{(2)}\right]$$

Followed by the normalisation process:

$$\tilde{S_t}^{(2)} = S_t^{(2)} - M_{2\times 12}\left(S_t^{(2)}\right)$$

- Step 8: estimating the seasonally adjusted series:

$$A_t^{(2)} = (C_t + I_t)^{(2)} = X_t^{(2)} - \tilde{S_t}^{(2)}$$

2.4.2 Advantages

The X11 decomposition includes many extra steps and features in order to overcome the drawbacks of classical decomposition. In particular, trend-cycle estimates are available for all observations including the end points, and the seasonal component is allowed to vary slowly over time. X11 also has some sophisticated methods for handling trading day variation, holiday effects and the effects of known predictors. It handles both additive and multiplicative decomposition. The process is entirely automatic and tends to be highly robust to outliers and level shifts in the time series.

2.5 SEATS Decomposition

SEATS stands for Seasonal Extraction in ARIMA Time Series, where the ARIMA model will be discussed later in this article. This procedure was developed at the Bank of Spain, and is now widely used by government agencies around the world. The procedure works only with quarterly and monthly data. So seasonality of other kinds, such as daily data, or hourly data, or weekly data, require an alternative approach.

For the algorithm of SEATS decomposition, `https://jdemetradocumentation.github.io/JDemetra-documentation/pages/theory/SA_SEATS.html` gives a brief introduction.

2.6 STL Decomposition

STL is a filtering procedure for decomposing a time series into trend, seasonal, and remainder components. STL has a simple design that consists of a sequence of applications of the LOESS smoother; the simplicity allows analysis of the properties of the procedure and allows fast computation, even for very long time series and large amounts of trend and seasonality smoothing.

2.6.1 Algorithms

The STL decomposition uses the LOESS smoother [4], which stands for the locally estimated scatter plot smoothing. The main idea of the LOESS smoothing is to divide the data in the whole time series into smaller clusters, and to fit a line or a curve to each cluster. The LOESS usually fits the local clusters with either a linear function or a quadratic one. Here we show an example of the linear function interpolation.

A LOESS smoother begins with raw data points, an example is shown in Figure 2. Note that at the same time value, there should be only one single observation of the value on the y-axis. That is to say, each data point should have a distinct x-value.

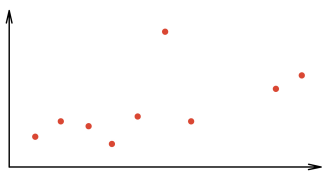


Figure 2: The raw data

To apply the LOESS smoother, we need to set the degree and the span, which are defined as follows:

- degree: the degree of the polynomial for local regression line
- span: the number of points in each local cluster

We denote the middle point of the i^{th} local cluster to be p_i and their surrounding points to be p_j. The distances between p_i and p_j are calculated and scaled to a $(-1, 1)$ range.

$$d(p_i, p_j) = |x_i - x_j|$$

$$d^*(p_j) = \frac{d(p_i, p_j)}{\max_k d(p_i, p_k)}$$

Then we conduct a weighted regression of same degree for all local clusters, the weight function is defined as follows [4]:

$$w(x) = \begin{cases} \left(1 - |x|^3\right)^3 & -1 < x < 1 \\ 0 & x \leq -1 \text{ or } x \geq 1 \end{cases}$$

The figure for the weight function is shown in Figure 3. We can easily notice that for some x close to zero, the weights are given very close to 1, with the middle points given the highest among all, while data points with farther distances are given really low weights. In this way, the several data points close to the centre of the local clusters are largely determining the regression results.

Note that if there are n data points, then this algorithm will produce n weighted regressions, one for each point.

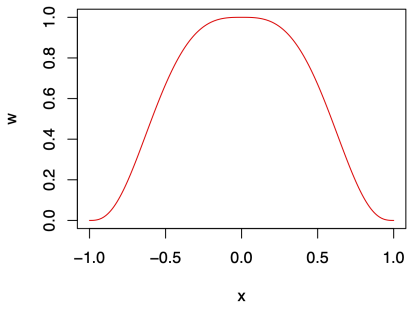
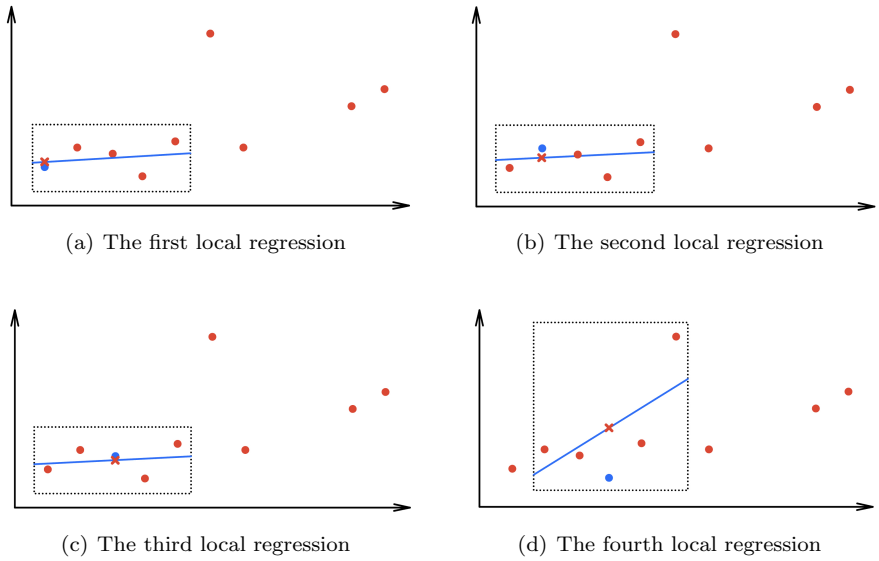


Figure 3: The weight function

There are iterative versions of LOESS and multivariate versions present [5], which will provide better smoothing power.

2.6.2 An example shown in figure

The following example in Figure 4 shows a simple LOESS smoothing process with degree=1 and span=5.



(a) The first local regression

(b) The second local regression

(c) The third local regression

(d) The fourth local regression

Figure 4: An example of a LOESS smoothing process

From above we will obtain the final smoothed curve as Figure 5.

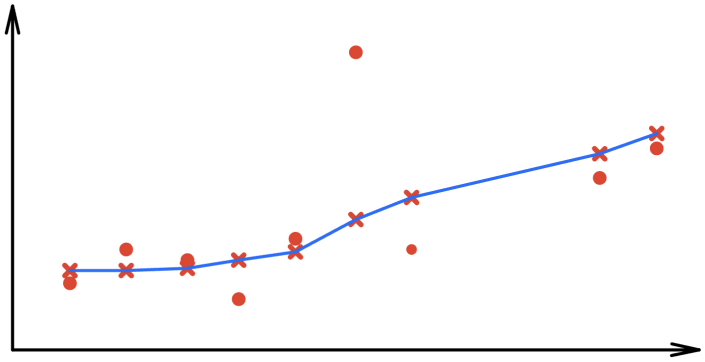


Figure 5: An example of the smoothed curve from LOESS smoother

3 Auto-Regressive Integrated Moving Average (ARIMA)

3.1 Components

The ARIMA model stands for Auto-regressive Integrated Moving Average, which includes the auto-regressive model (AR), the difference step (corresponding to the "I" in the model) and the moving average process (MA).

3.1.1 The auto-regressive model

The auto-regressive model specifies that the output variable depends linearly on its own previous values and a stochastic term, which is an imperfectly predictable term. The model is originally defined by a difference equation:

$$X_t = \sum_{i=1}^{p} \varphi_i X_{t-i} + \epsilon_t$$

Which can be simplified to:

$$X_t = c + \sum_{i=1}^{p} \phi_i y_{t-i} + \epsilon_t$$

Where φ_i and ϕ_i are parameters and ϵ_t is the white noise[1]. We often define this as the $AR(p)$ model, which is the auto-regressive model of order p.

3.1.2 Difference and stationarity

For a stationary time series, we require the mean value and variance of the y-value to be constant, and the covariance to be only related with the time interval of the series, but not the time t. Consider a non-stationary time series, we perform difference to it, that is:

$$Z_t = X_t - X_{t-1}$$

Or, using the difference operator:

$$Z_t = D^d X_t$$

Where d is the order of the difference process.

3.2 Determining the parameter values

The three parameters, p, d, q, gives the ARIMA(p, d, q) model.

[1] A series of independent normal distributions with mean 0 and variance σ^2.

3.2.1 Parameter *d*

We plot the time series (and the differenced ones if necessary) and perform the ADF test to check the stationarity [6], the parameter d is given by the order of the difference.

Consider the $AR(p)$ model:

$$X_t = \mu + \sum_{i=1}^{p} \varphi_i X_{t-i} + \epsilon_t$$

Subtracting X_{t-1} from both sides gives:

$$\Delta X_t = \mu + \delta X_{t-1} + \sum_{i=1}^{p} \beta_i \Delta X_{t-i} + \epsilon_t$$

The ADF test statistic is given by:

$$t_{\hat{\beta}_i} = \frac{\hat{\beta}_i}{SE\left(\hat{\beta}_i\right)}$$

The test is conducted under following assumptions:

- Null Hypothesis (H_0): There exists a unit root in the time series and it is non-stationary. Unit root $= 1$ or $\delta = 0$.
- Alternate Hypothesis (H_1): There exists no unit root in the time series and it is stationary. Unit root< 1 or $\delta < 0$.

Therefore we expect the p value to be less than the significance level (0.05), so that H_0 is rejected and we consider the series to be stationary.

Figure 6: Effect of difference

3.2.2 Parameter p

The parameter p is defined as the order of the AR model. We use the PACF plot to determine the suitable value for p.

The PACF, standing for Partial Auto-correlation Function, show the correlation between a time series and its lagged values, after accounting for the correlations at shorter lags [7]. This can be useful for identifying the most important lags to include in a time series model, as well as for identifying the presence of seasonal patterns in the data. It is defined as:

$$PACF\left(k\right) = \frac{\mathbb{E}\left(Z_t - \mathbb{E}\left(Z_t\right)\right)\left(Z_{t-k} - \mathbb{E}\left(Z_{t-k}\right)\right)}{\sqrt{\mathbb{E}\left(Z_t - \mathbb{E}\left(Z_t\right)\right)^2}}$$

Simplified to:

$$PACF\left(k\right) = \frac{Cov\left[\left(Z_t - \bar{Z}_t\right),\ \left(Z_{t-k} - \bar{Z_{t-k}}\right)\right]}{\sqrt{Var\left(Z_t - \bar{Z}_t\right)}\sqrt{Var\left(Z_{t-k} - \bar{Z_{t-k}}\right)}}$$

3.2.3 Parameter q

The parameter q is defined as the order of the MA model. We use the ACF plot to determine the suitable value for q.

ACF plots show the correlation between a time series and lagged versions of itself [7]. The ACF plot can be used to identify the number of lags that are needed for a time series model. For example, if the ACF plot shows a strong correlation between the time series and its lag-1, lag-2, and lag-3 values, then a time series model that includes these lags would likely be a good fit [7]. Auto-correlation Function (ACF) is defined as:

$$ACF\left(k\right) = \sum_{t=k+1}^{n} \frac{\left(Z_t - \bar{Z}\right)\left(Z_{t-k} - \bar{Z}\right)}{\sum_{t=1}^{n}\left(Z_t - \bar{Z}\right)^2}$$

The following shows an example of an ACF and a PACF plot:

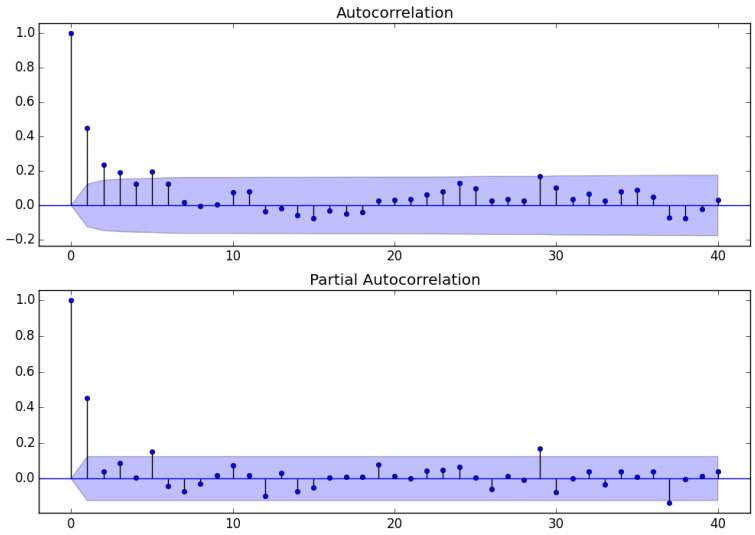


Figure 7: An example of an ACF and a PACF plot

In this example, both the ACF plot and the PACF plot show a large cut-off at lag 2, therefore we take the $AR(2)$ and the $MA(2)$ model.

However, using plots to determine the parameters p and q may bring errors due to subjective reasons, another alternative is to use the AIC and BIC values, defined as:

$$AIC = -2\ln(L) + 2K$$

$$BIC = -2\ln(L) + K\ln(n)$$

Where L denotes the maximum likelihood of the model, K denotes the number of parameters in the model, and n denotes the sample size. Note that when n gets very large, we use the BIC Bayesian criteria.

The values of p and q that give the combination of the smallest AIC and BIC values are chosen for the model. Hence we construct the complete ARIMA(p, d, q) model.

References

[1] Rob J Hyndman and Australia George Athanasopoulos Monash University. *6.3 Classical decomposition*. URL: `https://otexts.com/fpp2/classical-decomposition.html` (visited on 08/07/2024).

[2] Dominique Ladiray and Benoit Quenneville. *Seasonal adjustment with the X-11 method*. Vol. 158. Springer Science & Business Media, 2012.

[3] Christian Harhoff. *Seasonal Adjustment*. Statistics Denmark, 2005.

[4] Robert B Cleveland et al. "STL: A seasonal-trend decomposition". In: *J. off. Stat* 6.1 (1990), pp. 3–73.

[5] William S Cleveland. "Robust locally weighted regression and smoothing scatterplots". In: *Journal of the American statistical association* 74.368 (1979), pp. 829–836.

[6] J Humberto Lopez. "The power of the ADF test". In: *Economics Letters* 57.1 (1997), pp. 5–10.

[7] Usman Abbas Yakubu and Moch Panji Agung Saputra. "Time series model analysis using autocorrelation function (ACF) and partial autocorrelation function (PACF) for E-wallet transactions during a pandemic". In: *International Journal of Global Operations Research* 3.3 (2022), pp. 80–85.