Time Series Forecasting in Python

Stephen CUI

December 12, 2023

Contents

Part I

Time waits for no one

Chapter 1

Understanding time series forecasting

1.1 Forecasting the historical mean

1.1.1 Implementing the historical mean baseline

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{A_i - F_i}{A_i}\right| \times 100 \tag{1.1}$$

A_i is the actual value at point i in time, and F_i is the forecast value at point i in time; n is simply the number of forecasts.

Chapter 2

Going on a random walk

2.1 Identifying a random walk

Definition 1 *A random walk is a series whose first difference is stationary and uncorrelated. This means that the process moves completely at random.*

2.1.1 Stationarity

A stationary time series is one whose statistical properties do not change over time. In other words, it has a constant mean, variance, and autocorrelation, and these properties are independent of time.

Augmented Dickey-Fuller (ADF) test

The augmented Dickey-Fuller (ADF) test helps us determine if a time series is stationary by testing for the presence of a unit root. If a unit root is present, the time series is not stationary. The null hypothesis states that a unit root is present, meaning that our time series is not stationary.

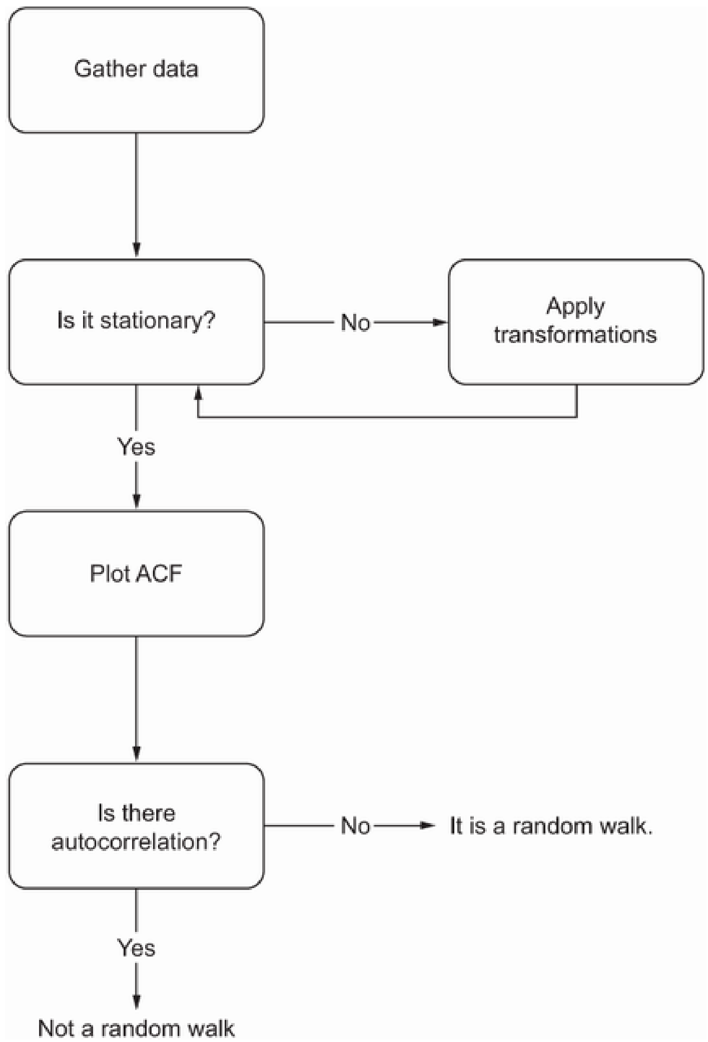

Figure 2.1: Steps to follow to identify whether time series data can be approximated as a random walk or not. The first step is naturally to gather the data. Then we test for stationarity. If it is not stationary, we apply transformations until stationarity is achieved. Then we can plot the autocorrelation function (ACF). If there is no autocorrelation, we have a random walk.

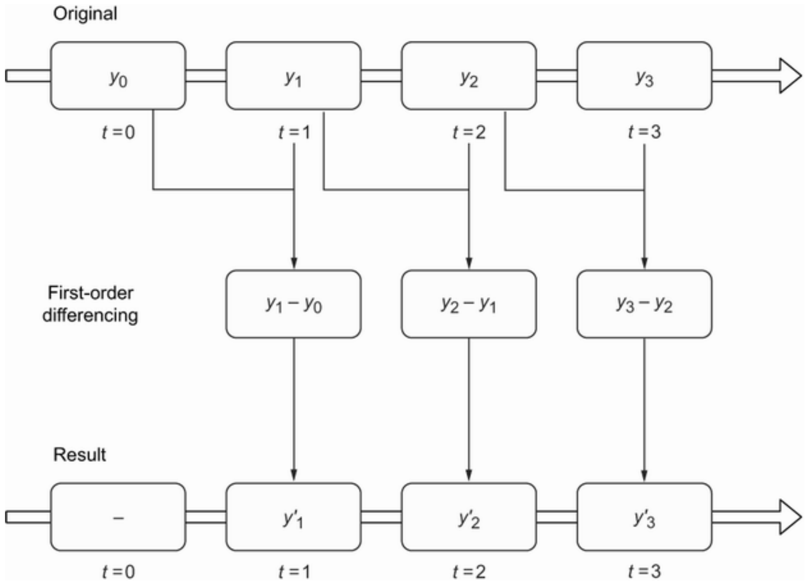

Figure 2.2: Visualizing the differencing transformation. Here, a first-order differencing is applied. Notice how we lose one data point after this transformation because the initial point in time cannot be differenced with previous values since they do not exist.

2.2 Forecasting a random walk

2.2.1 Forecasting on a long horizon

You should be convinced that forecasting a random walk on a long horizon does not make sense. Since the future value is dependent on the past value plus a random number, the randomness portion is magnified in a long horizon where many random numbers are added over the course of many timesteps.

Part II

Forecasting with statistical models

Chapter 3

Modeling a moving average process

3.1 Defining a moving average process

A **moving average process**, or the moving average (MA) model, states that the current value is linearly dependent on the current and past error terms. The error terms are assumed to be mutually independent and normally distributed, just like white noise.

A moving average model is denoted as $MA(q)$, where q is the order. The model expresses the present value as a linear combination of the mean of the series μ, the present error term ε_t, and past error terms ε_{t-q}. The magnitude of the impact of past errors on the present value is quantified using a coefficient denoted as θ_q. Mathematically, we express a general moving average process of order q as in Equation 3.1.

$$y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} \tag{3.1}$$

3.2 Forecasting a moving average process

对于预测范围，移动平均模型具有特殊性。 MA(q) 模型不允许我们一次性预测未来多步。请记住，移动平均模型线性依赖于过去的误差项，并且这些项在数据集中未观察到 - 因此必须递归估计它们。这意味着对于 MA(q) 模型，我们只能预测未来的 q 步。超过该点所做的任何预测都不会包含过去的误差项，并且模型只会预测平均值。因此，对未来超过 q 步的预测没有附加值，因为预测将持平，因为仅返回平均值，这相当于基线模型。

Forecasting using the MA(q) model

When using an MA(q) model, forecasting beyond q steps into the future will simply return the mean, because there are no error terms to estimate beyond q steps. We can use rolling forecasts to predict up to q steps at a time in order avoid predicting only the mean of the series.

Summary

1. A moving average process states that the present value is linearly dependent on the mean, present error term, and past error terms. The error terms are normally distributed.
2. You can identify the order q of a stationary moving average process by studying the ACF plot. The coefficients are significant up until lag q only.
3. You can predict up to q steps into the future because the error terms are not observed in the data and must be recursively estimated.
4. **Predicting beyond q steps into the future will simply return the mean of the series. To avoid that, you can apply rolling forecasts.**
5. If you apply a transformation to the data, you must undo it to bring your predictions back to the original scale of the data.
6. The moving average model assumes the data is stationary. Therefore, you can only use this model on stationary data.

Chapter 4

Modeling an autoregressive process

Chapter 5

Modeling complex time series

5.1 Identifying a stationary ARMA process

Identifying a stationary ARMA process

If your process is stationary and both the ACF and PACF plots show a decaying or sinusoidal pattern, then it is a stationary ARMA(p,q) process.

5.2 Devising a general modeling procedure

We saw that if both the ACF and PACF plots display a sinusoidal or decaying pattern, our time series can be modeled by an ARMA(p,q) process. However, neither plot was useful for determining the orders p and q. With our simulated ARMA(1,1) process, we noticed that coefficients were significant after lag 1 in both plots.

Therefore, we must devise a procedure that allows us to find the orders p and q. This procedure will have the advantage that it can also be applied in situations where our time series is non-stationary and has seasonal effects. Furthermore, it will also be suitable for cases where p or q are equal to 0, meaning that we can move away from plotting the ACF and PACF and rely entirely on a model selection criterion and residual analysis. The steps are shown in Figure 5.1.

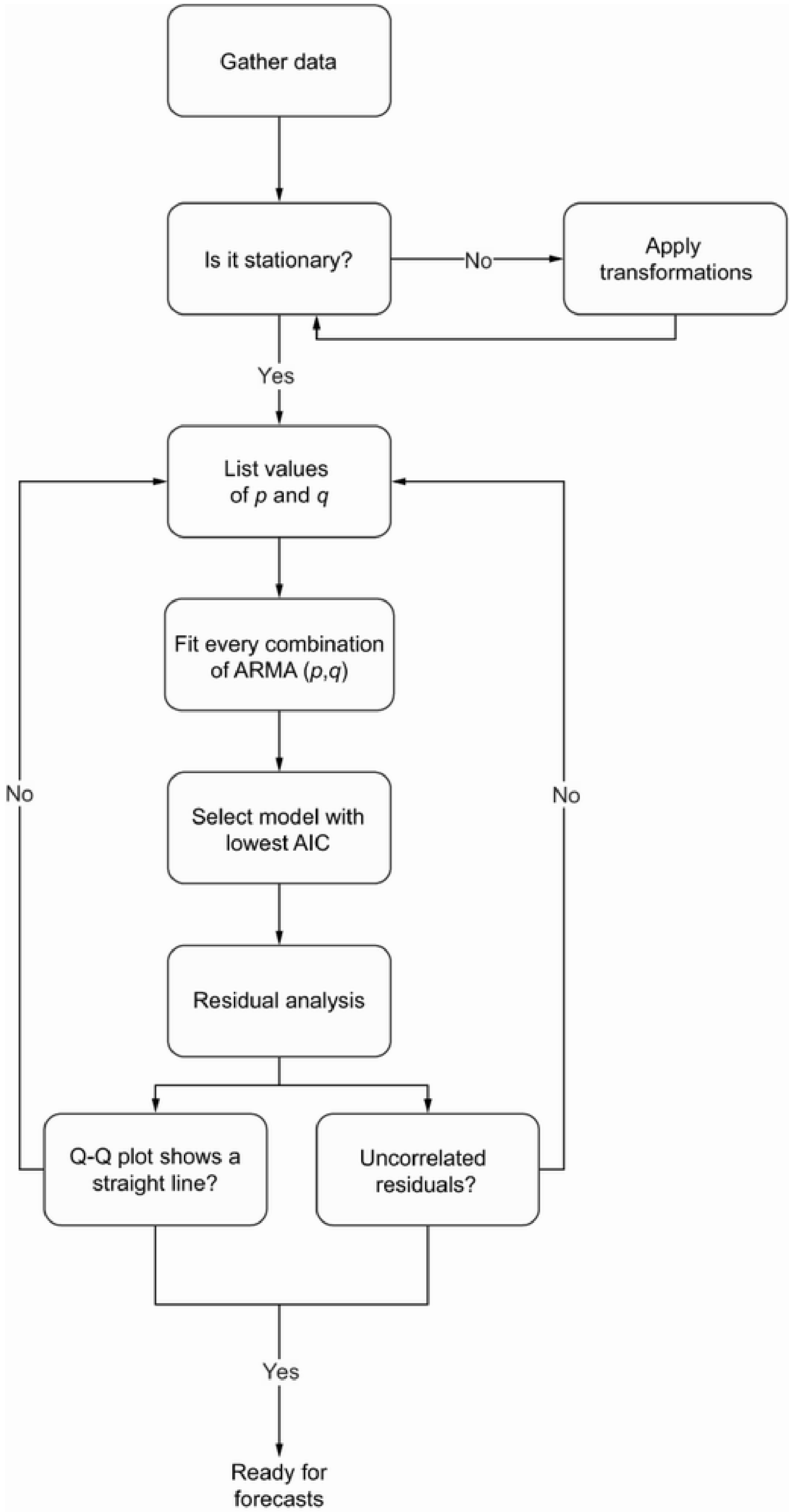

Figure 5.1: General modeling procedure for an ARMA(p,q) process. The first steps are to gather the data, test for stationarity, and apply transformations accordingly. Then we define a list of possible values for p and q. We then fit every combination of ARMA(p,q) to our data and select the model with the lowest AIC. Then we perform the residual analysis by looking at the Q-Q plot and the residual correlogram. If they approach that of white noise, the model can be used for forecasts. Otherwise, we must try different values for p and q.

5.2.1 Understanding the Akaike information criterion (AIC)

The AIC estimates the quality of a model relative to other models. Given that there will be some information lost when a model is fitted to the data, the AIC quantifies the relative amount of information lost by the model. The less information lost, the lower the AIC value and the better the model.

The AIC is a function of the number of estimated parameters k and the maximum value of the likelihood function for the model $\hat{L}$

$$AIC = 2k - 2\ln(\hat{L}) \tag{5.1}$$

You can see how fitting a more complex model can penalize the AIC score: as the order (p,q) increases, the number of parameters k increases, and so the AIC increases.

The likelihood function measures the goodness of fit of a model. It can be viewed as the opposite of the distribution function. Given a model with fixed parameters, the distribution function will measure the probability of observing a data point. The likelihood function flips the logic. **Given a set of observed data, it will estimate how likely it is that different model parameters will generate the observed data**.

5.2.2 Understanding residual analysis

QUANTITATIVE ANALYSIS: APPLYING THE LJUNG-BOX TEST

The Ljung-Box test is a statistical test that determines whether the autocorrelation of a group of data is significantly different from 0.

In time series forecasting, we apply the Ljung-Box test on the model's residuals to test whether they are similar to white noise. The null hypothesis states that the data is independently distributed, meaning that there is no autocorrelation. If the p-value is larger than 0.05, we cannot reject the null hypothesis, meaning that the residuals are independently distributed. Therefore, there is no autocorrelation, the residuals are similar to white noise, and the model can be used for forecasting. If the p-value is less than 0.05, we reject the null hypothesis, meaning that our residuals are not independently distributed and are correlated. The model cannot be used for forecasting.

Summary

- The autoregressive moving average model, denoted as ARMA(p,q), is the combination of the autoregressive model AR(p) and the moving average model MA(q).

- An ARMA(p,q) process will display a decaying pattern or a sinusoidal pattern on both the ACF and PACF plots. Therefore, they cannot be used to estimate the orders p and q.
- The general modeling procedure does not rely on the ACF and PACF plots. Instead, we fit many ARMA(p,q) models and perform model selection and residual analysis.
- Model selection is done with the Akaike information criterion (AIC). It quantifies the information loss of a model, and it is related to the number of parameters in a model and its goodness of fit. The lower the AIC, the better the model.
- The AIC is relative measure of quality. It returns the best model among other models. For an absolute measure of quality, we perform residual analysis.
- Residuals of a good model must approximate white noise, meaning that they must be uncorrelated, normally distributed, and independent.
- The Q-Q plot is a graphical tool for comparing two distributions. We use it to compare the distribution of the residuals against a theoretical normal distribution. If the plot shows a straight line that lies on y = x, then both distributions are similar. Otherwise, it means that the residuals are not normally distributed.
- The Ljung-Box test allows us to determine whether the residuals are correlated or not. The null hypothesis states that the data is independently distributed and uncorrelated. If the returned p-values are larger than 0.05, we cannot reject the null hypothesis, meaning that the residuals are uncorrelated, just like white noise.

Part III

Large-scale forecasting with deep learning

Part IV

Automating forecasting at scale