Time Series Modeling
Linear Trends, Stationarity, and Autoregression Models (TS2)

1 Introduction

We discuss a class of time series models, analogous to generalized linear models (including the subclass of linear regression models), that is a function of a linear combination of coefficients on predictor variables. For time series models, the predictors are time-dependent events preceding the event of interest. The predictor coefficients thus form a linear combination of polynomials. Polynomials are functional relationships of predictors with coefficients for which only the operations of addition, subtraction, multiplication, and non-negative integer exponents apply. These relationships have well behaved mathematical properties desirable in time series models. The exponents applied to the predictors are time-based backward shifts of the series. The coefficients are a representation of the weighted effect on the event of interest. We now discuss some of the details of linear modeling of time series.

2 Stationarity and Invertibility Conditions for a Linear Process

When we construct a linear regression model, we test for conformance to a set of assumptions, viz., the Gauss-Markov assumptions. Time series models also require a set of assumptions to achieve an adequate, reliable forecast model. Just as with linear regression models which require conditions of constant variance for a stable inverted design matrix, time series models require a stationary process and the ability to invert a linearly-related set of polynomials.

2.1 Stationarity

Time series modeling is about analyzing the autocovariance among the relevant realizations of a finite set of random variables indexed, for this course but not a necessity, by fixed time intervals. A condition for constructing an adequate linear regression model is a response with a fixed mean and constant variance. For an adequate time series model, we also require a fixed mean and constant variance for any given time interval in the series of interest. Fixed mean and constant variance in time series modeling is known as stationarity. However, there are flavors of stationarity we must consider.

When a time series has a constant mean, finite variance, and the covariance is dependent only on the lag (time interval from the time of the event of interest), we have what is known as weak stationarity (also known as covariance stationarity, stationarity in the weak sense, and second-order stationarity). We will use the term stationarity to mean weak stationarity unless otherwise stated.

An important special case of stationarity is known as strict stationarity. This type of stationarity has that, for any two equal-length time spans from a time series realization, the two spans exhibit identical statistical characteristics; i.e., the two spans have equal means and equal variances. The implication is that the two spans have the same probability distributions and the same covariance.

Hence, a strictly stationary time series with finite variance is stationary, but the converse is not necessarily true: a stationary time series is not necessarily strictly stationary.

When a time series follows a normal distribution, the series is strictly stationary. This is true as normal distributions have fixed means, finite variance, the mean and variance are independent of each other, and each realization is drawn from an identical distribution. Note the similarity to the linear regression Gauss-Markov assumptions. The difference between the time series strict stationarity condition and the Gauss-Markov assumptions is the time series is not required to have independent observations.

2.2 Linear Trend

Stationarity is a critical characteristic for time series model construction and the demonstration that a constructed time series model is an adequate representation of the data. If a time series is not stationary, it is often possible to transform the series to achieve this condition. Consider a stationary time series, $X_t,\ t = 0, \pm 1, \pm 2, \ldots$, superposed on a linear, upward or downward trend:

$$Y_t = X_t + at + b, \tag{1}$$

where a is the slope of the linear time t trend line and b is the linear trend line zero intercept. Clearly, a time series with a linear trend violates the stationarity condition required to construct a viable time series forecast model. However, if we take a first difference of Y_t, we obtain

$$\begin{aligned}
\bigtriangledown Y_t = Y_t - Y_{t-1} &= (X_t + at + b) - (X_{t-1} + a(t-1) + b) \\
&= X_t - X_{t-1} + a(t - (t-1)) + (b - b) \\
&= \bigtriangledown X_t + a \\
\implies \bigtriangledown Y_t - a &= \bigtriangledown X_t, \qquad (2)
\end{aligned}$$

which indicates the first difference (rate) of the time series process Y_t is a stationary process.

There are other transformations that may be applied to a non stationary time series that results in a stationary series. We shall encounter such transformations throughout this course.

3 Autoregressive Models (AR)

We consider a class of time series models that require the time dependent process to be stationary. This class of models may be defined in terms of linear difference equations (not necessarily the first difference from above) with constant coefficients. This linear difference of constant coefficients defines a parametric family of stationary processes one member being AutoRegressive (AR) models. The linear structure of AR models allows for a simple theory of linear forecasting.

To understand AR models, we first must revisit white noise.

3.1 White Noise (WN)

A time series with mean zero and finite, constant variance with each realization through time independent of any other realization is called white noise. White noise is denoted $Z_t \sim \mathcal{WN}(0, \sigma^2)$, where Z_t is a time-ordered sequence of identically distributed random variables with zero mean and finite variance ($\sigma^2 < \infty$). If the white noise realizations are each independent in time, then we say the white noise is identically independently distributed ($Z_t \sim \text{iid } \mathcal{WN}(0, \sigma^2)$).

White noise processes are important to time series model construction as we can generate a large class of time series processes using white noise as a basis.

3.2 AR Structure

Suppose we have a mean zero time series $X_t,\ t = 0, \pm 1, \pm 2, \ldots$ This process is said to be ARMA(p) if X_t is stationary and if, for all t,

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t, \tag{3}$$

where $Z_t \sim \mathcal{WN}(0, \sigma^2)$. It is clear that the process X_t is generates Z_t, a white noise series. A compact way to write Equation 3 is

$$\phi(B) X_t = Z_t, \quad t = 0, \pm 1, \pm 2, \ldots, \tag{4}$$

where ϕ is the respective pth polynomial

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p. \tag{5}$$

B is the back shift operator

$$B^j X_t = X_{t-j},\ j = 0, \pm 1, \pm 2, \ldots \tag{6}$$

The coefficients $\phi_k,\ k = 1, 2, \ldots, p$ are the autoregressive (AR) polynomial coefficients. Hence,

$$\phi(B) X_t = Z_t, \tag{7}$$

is an autoregressive (AR) process of order p and is denoted as $AR(p)$. Notice that if the coefficients ϕ are correctly chosen, the AR model is a white noise generator.

3.3 AutoCorrelation Function (ACF) and Partial AutoCorrelation Function (PACF)

An important concern with many time series is the bivariate correlation structure between time and the series of realizations. As time is the index set sequencing the series, the correlation structure is an autocorrelation. Autocorrelation measures the dependency of preceding events on successive events. Hence, is there a linear combination of realizations that determines the level of a successive realization within random error? The coefficients of the linear combination reflects the level of each autocorrelation value at each previous event or lag. The autocorrelation function, ACF, gives an assessment of the order of the linear polynomial that adequately describes the autocorrelation. It gives the number of lags prior to the current event needed for an unbiased forecast of its value.

The sign of the ACF values determines whether the lag lies on the same side of the mean, or alternates across the mean. For example, a positive autocorrelation at lag 1, say, reflects a tendency for successive realizations to lie on the same side of the mean, whereas a negative sign on the value at lag 1 reflects a tendency for successive observations to lie on opposite sides of the mean.

The sample ACF can be determined for any time series and is not restricted to stationary processes. If a series contains a trend, the ACF will show a slow decay. If a series contains a periodic pattern, the ACF will exhibit the same cyclicality. We see, then, that the ACF may be used as an indicator of nonstationarity.

The Partial AutoCorrelation Function (PACF), like the ACF, gives information on a time series' interdependencies. Unlike the ACF, the PACF at, say lag k, is the correlation between X_t and $X_{t-k},\ k = 0, 1, 2, \ldots$, adjusted for the intervening realizations $X_{t-(k-1)}, X_{t-(k-2)}, \ldots, X_{t-1}$. The PACF is the correlation of X_t and X_{t-k} after regressing on the intervening realizations.

The PACF provides information on the order of an AR model's autoregressive order; i.e., the $AR(p)$. Figure 1 is an ACF of the output of a radio telescope receiver after the signal was subjected to first differencing. The only significant lag of the ACF is lag 1, and suggests the differenced time series is stationary. Figure 2 is a PACF of the output of the same radio telescope receiver after the signal was subjected to first differencing. The significant lags shown in the PACF are lags 1 through 4. This suggests an AR(4) model may adequately represent the time series (first difference) process.

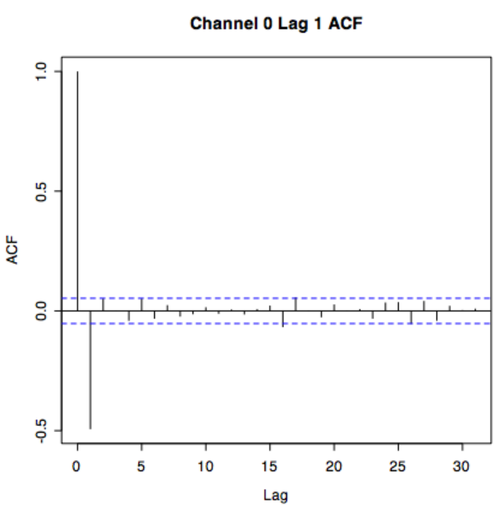

Figure 1: The autocorrelation function (ACF) plot a radio telescope receiver signal after a first difference transformation.

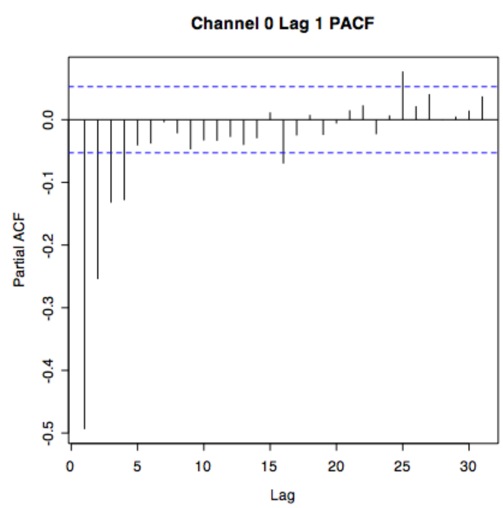

Figure 2: The partial autocorrelation function (PACF) plot of a radio telescope receiver signal after a first difference transformation.

4 Unit Root Nonstationarity)

Consider the AR(p) model

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t. \tag{8}$$

We assess whether this time series model is stationary by examining its characteristic equation and the concomitant roots. The roots are the solutions to the characteristic equation. We find the solutions to Equation 8 by rewriting it as as a lag polynomial:

$$(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p) y_t = \epsilon_t. \tag{9}$$

Now we replace the lag operator B with a variable, say z, and set this to zero as

$$(1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p) y_t = 0, \tag{10}$$

which gives us the characteristic equation whose roots are the values of z are the solutions to this equation. There are p values of z and some of these values may not be unique. The time series y_t is stationary if all of the roots lie outside the unit circle.

Some of these roots may be complex numbers which may suggest cycling in the time series (see Section 5 below). If all the roots are real numbers, i.e., the roots are not complex numbers, then we can say that y_t is stationary if the absolute values of all of these real roots are greater than one. If a root equals ± 1, it is a unit root. If there is at least one unit root, or if any of the roots lie between ± 1, then the series is not stationary.

Suppose as an example we have that Equation 8 is

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t, \tag{11}$$

so the characteristic equation is

$$1 - \phi_1 z - \phi_2 z^2 = 0. \tag{12}$$

The roots are the solution to the familiar quadratic equation:

$$z = \frac{-(-\phi_1) \pm \sqrt{(-\phi_1)^2 - 4(-\phi_2)(1)}}{2\phi_2} = \frac{\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{2\phi_2}. \tag{13}$$

Recall that the quadratic equation gives two solutions and hence two roots: $\pm z$. If the roots are real and the absolute value of z is ≥ 1, then the time series is stationary. if $\mid z \mid < 1$, then the time series is not stationary.

5 Time Series Cycles (Business Cycles)

Data associated with time, particularly data with sinusoidal patterns, have characteristics that are modeled in the complex plane; i.e., the models use complex numbers. The class of such models lies within the realm of complex analysis which uses real numbers combined with imaginary numbers. An additional diagnostic for candidate model residuals is to look for unit roots which, if they exist, suggest the residuals are not stationary. Nonstationary residuals indicate the candidate model may not adequately represent the time series.

Recall that the solution to $\sqrt{4} = \pm 2$, i.e., there are two solutions. We also know that when we use only real numbers as there is no solution to $\sqrt{-4}$. However, if we define $i = \sqrt{-1} \Rightarrow i^2 = -1$, then we may represent $i^2 = (0 + i1)(0 + i1) = 0^2 + (2 \cdot 0 \cdot i1) + i^2 \cdot 1^2 = -1$. Here, 0 is the real part of the complex number and 1 is the complex part of the complex number which we decompose as $z = x + iy$. The real part of z is $\mathrm{Re}(z) = x$ and the imaginary part is $\mathrm{Im}(z) = y$.

Figure 3 is a plot of z in the complex plane when z is a line. Figure 4 is a plot of z when z is a sinusoid. The sinusoid plot is iy versus x where x is time. Suppose we have a set of data which manifest as a period time series. Let w be a time series of the form

$$w = c_0 + c_1 z + c_2 z^2 + \cdots + c_n z^n, \quad n = 0, 1, 2, ... \tag{14}$$

We represent the roots of w as

$$w = \sqrt[n]{z} = \sqrt[n]{r}\left(\cos\frac{\theta + 2k\pi}{n} + i\sin\frac{\theta + 2k\pi}{n}\right), \quad k = 0, 1, \ldots, n-1. \tag{15}$$

These n roots lie on a circle of radius $\sqrt[3]{r}$ with center a the origin and constitute the vertices of a regular polygon of n sides. For example, the square root of $w = \sqrt{z}$ has two roots which are

$$z_1 = \sqrt{r}\left(\cos\frac{\theta}{2} + i\sin\frac{\theta}{2}\right) \tag{16}$$

and

$$z_2 = \sqrt{r}\left[\cos\left(\frac{\theta}{2} + \pi\right) + i\sin\left(\frac{\theta}{2} + \pi\right)\right] = -z_1 \tag{17}$$

which lie symmetrically about the origin on the x-axis. In particular, suppose

$$\sqrt{4i} = \pm 2\left(\cos\frac{\pi}{4} + i\sin\frac{\pi}{4}\right) = \pm(\sqrt{2} + i\sqrt{2}). \tag{18}$$

The above is the simplified mathematical framework for why we look for business cycles using complex roots.

Another way to see these (business) cycles is to consider a circle in the x,y-plane centered at (0,0) with unit radius; i.e., $r = 1$. Then, by the Pythagorean Theorem,

$$\begin{aligned} r^2 &= x^2 + y^2 \\ &= 1 \quad \text{for a unit circle.} \end{aligned} \tag{19}$$

From trigonometry, for $x = \cos t$ and $y = \sin t$, we have

$$\begin{aligned} r^2 &= x^2 + y^2 \\ &= (\cos t)^2 + (\sin t)^2 \\ &= \cos^2 t + \sin^2 t \\ &= 1 \end{aligned} \tag{20}$$

Recall that a solution to a quadratic equation is through the use of the quadratic formula

$$\frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \tag{21}$$

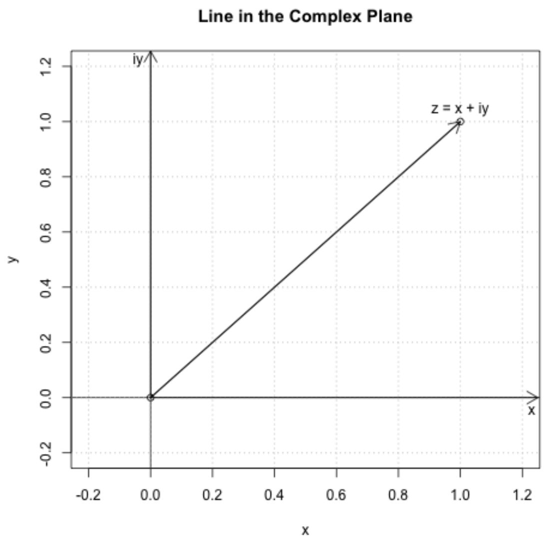

Figure 3: A vector line in the complex plane.

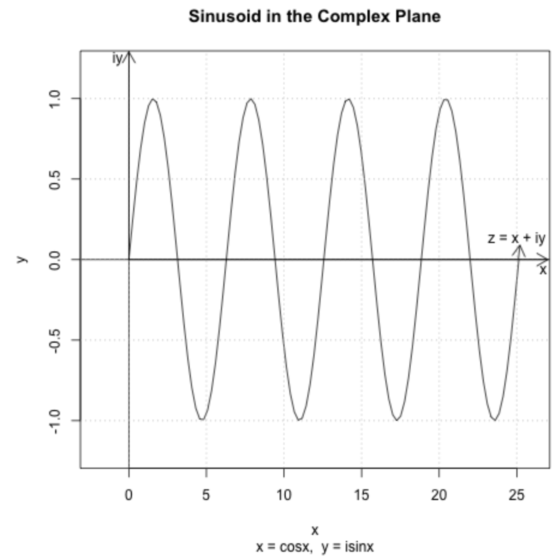

Figure 4: A sinusoidal time series in the complex plane.

which yields expressions for the two solutions to $ax^2+bx+c=0$. However, this formula may involve taking the square root of negative numbers. We may treat these solutions as ordinary numbers if we utilize $\sqrt{-1}\cdot\sqrt{-1}=-1$ such that $i^2=-1$, where i denotes accessing the imaginary numbers in the complex plane. Think of the complex plane in which the x-axis is the real number line and the y-axis is the imaginary number line.

Now let a cyclical function be such that $x=\cos t$ and $y=i\sin t$. Then, using de Moivre's Formula, we have that

$$z=r(\cos t+i\sin t) \tag{22}$$

where z is from the complex numbers. As we know $r=1$ for the unit circle, then from Equation 22, we obtain by squaring z,

$$\begin{aligned} z^2 &= r^2(\cos(t+t)+i\sin(t+t)) \\ &= \cos 2t + i\sin 2t \\ &= \cos^2 t - \sin^2 t + i(2\sin t\cos t) \\ &= \cos^2 t + \sin^2 t - i2\sin t\cos t \\ &= 1 - iat \end{aligned} \tag{23}$$

$$\implies z = \sqrt{1-iat} \tag{24}$$

Thus, we have at complex roots about the unit circle, and, for $a\neq 0$ we have evidence of cycling. This result applies to cycles in time series data which we often refer to as seasonality.

The null hypothesis of the Augmented Dickey-Fuller test is that at least one unit root is present which is interpreted as the process is non-stationary.

The business cycles are calculated for an AR(3) assuming $\phi_1 = 0.4386$, $\phi_2 = 0.2063$, $\phi_3 = -0.1559$, the roots (complex) are $1.6161 + 0.8642i$, $-1.0902 - 0i$, and $1.6161 - 0.88642i$. The moduli of the complex roots are, for Re(root) = real component and Im(root) = the imaginary part:

$$\begin{aligned} \text{mod } (1.6161 + 0.8642) &= |Re(1.6161 + 0.8642i) + Im(1.6161 + 0.8642i)| \\ &= \sqrt{1.6161^2 + 0.8642i^2} \\ &= \sqrt{2.611779 + 0.7468(1)} \ \text{ as } (\sqrt{-1})^2 = (-1)^2 = 1 \\ &= 1.8326 \end{aligned} \tag{25}$$

Repeat for each complex root.

To calculate the business cycles, use Tsay (2010), p. 42, as:

$$\begin{aligned} k &= \frac{2\pi}{\cos^{-1}\left(\frac{\phi_1}{2\sqrt{-\phi_2}}\right)} \\ &= \frac{2\pi}{\cos^{-1}\left(\frac{1.6161}{2\sqrt{1.8326}}\right)} \\ &\approx 12.7952 \end{aligned} \tag{26}$$

Similarly for the other complex roots.

6 Model Comparison

The norm in constructing time series models in to try more than one. This is due to there being no "true" model for a given time series, only approximations that converge toward a true model. We must thereby determine which model is "best". Ultimately, the best model is the model with the most accurate and precise forecasts. Unless the time period of the time series is short, it may take too long to know which model has the most accurate and precise forecasts. As such we use model comparison methods that are independent of time to make forecasts. We discussion methods below: The Akaike Information Criterion (AIC), in-sample comparison, and out-of-sample comparison. These methods do not assess the quality of a single model, they are for model comparisons only.

6.1 Akaike Information Criterion (AIC)

The Akaike Information Criterion (Akaike, 1973) is defined as

$$\text{AIC} = \frac{-2}{n} \times \ln(\hat{\sigma}^2_\epsilon) + \frac{2}{n} \times p, \tag{27}$$

where n is the length of the time series, σ^2_ϵ is the maximum likelihood estimate of the model error variance, and p is the order of the AR model. As we shall see, p will be included with additional model parameters in the upcoming modules. The second term is a penalty term with lower values resulting in lower and hence better AIC outcomes. The purpose of the penalty term is to reward for parameter parsimony. When comparing models, the lowest AIC suggests the best model. This is not a guarantee of the lowest AIC model being the best, however.

Another related comparison method is the Bayesian Information Criterion (BIC). Compared to the AIC, the BIC selects a lower AR model for larger n.

6.2 In-Sample Comparison

Recall from general linear modeling (regression, ANOVA), we can choose between constructing an effects model or a prediction model. The predictive model is its namesake: prediction. The effects model's purpose is to understand explanatory variable behavior on the response variable. The effects analog in time series models is to understand historical behavior. To test the viability of the model estimates for understanding the time series structure, we use in-sampling comparison among candidate models.

In-sample method use all the data in the time series to construct a model and make comparisons. In-sample comparison uses model diagnostics, AIC, and BIC for model comparison. Hence, the outcomes of the various model diagnostic tests and plots are used to assess best fit of a model. These diagnostics and the lowest AIC values aid in choosing a best model.

6.3 Out-of-Sample Comparison

When using a time series model for forecasting, we use forecasting performance to compare models. One measure of forecast performance is the mean square forecast error (MSFE) of out-of-sample forecasts. Obtaining this measure is also known as backtesting. Backtesting is a form of cross-validation. As the time series is divided into a training sample and an evaluation sample, backtesting uses one or more step ahead forecasts.

The backtesting procedure is as follows:

1. Choose the time at which to separate the time series into a training series $x_t,\ t = 1, 2, \ldots, h$ and an evaluation series $x_{h+1}, x_{h+2}, \ldots, x_n,\ n$ is the time series length. The evaluation series is used to calculate the MSFE so the time span must be long enough to permit prudent accuracy of the MSFE.

2. Construct a time series model using the training series. Obtain a one step (or more) ahead forecast. Calculate the forecast error as the difference between the model forecast and the corresponding realization from the evaluation series: $\hat{\eta}_h(k) = x_{h+k} - \hat{x}_h,\ k = 1, 2, \ldots$ are number of steps ahead desired.

3. Increase the training data series by k values such that $x_t,\ t = 1, 2, \ldots, h+k$ and the evaluation series is $x_{h+k+1}, x_{h+k+2}, \ldots, x_n$. Re-estimate the model from the new training set. Calculate the forecast error as the difference between the new model forecast and the corresponding realization from the evaluation series: $\hat{\eta}_{h+k}(k) = x_{h+k+1} - \hat{x}_{h+k}$.

4. Repeat step 3 until the evaluation series is absorbed into the training series such that the forecast error is $\hat{\eta}_{h+k}(k) = x_n - \hat{x}_{n-1}$.

The MSFE is determined as

$$\text{MSFE}_m = \frac{\sum_{j=h}^{n-1} (\hat{\eta}_j(k))^2}{n-h}, \tag{28}$$

where $m = 1, 2, \ldots$ is an index representing the model subjected to backtesting. We then choose the model m with the lowest MSFE or, more commonly, the square root of MSFE.

Two other commonly used forecasting performance measures, there are several, are the mean absolute forecast error and bias:

$$\text{MAFE}_m = \frac{\sum_{j=h}^{n-1} \mid \hat{\eta}_j(k) \mid}{n-h} \text{ and } \text{Bias}_m = \frac{\sum_{j=h}^{n-1} \hat{\eta}_j(k)}{n-h}. \tag{29}$$

6.4 Model Averaging

There are occasions when two or more models are retained either by choice or because the models all forecast without bias (the expected forecast error of each model is zero). We then may combine the multiple model forecasts for forecast averaging. Let $\hat{x}_{i,h+k}$ be the k-step ahead forecast of model $i,\ i = 1, 2, \ldots, m$ at the forecast model origin h. Then the forecast average is

$$\hat{x}_{h+1} = \sum_{i=1}^{m} w_i \hat{x}_{i,h+k}, \tag{30}$$

where w_i is the nonnegative real valued weight of model i for which $\sum_{i=1}^{m} w_i = 1$. When $w_i = 1/m$, we have the simple average forecast.

7 Remarks

We have seen from the above sections how the model-building process involves EDA, model selection, model diagnostics, and methods for evaluating the viability of the model forecasts. It may seem complex and involved at this stage, however, over the next several weeks, applications of the various techniques will become familiar. Each of the methods and techniques allow for the construction of useful and viable forecasting time series models.

References

Ruey S. Tsay. *Analysis of financial time series.* Wiley, Cambridge, MA, 2010.

H. Akaike. Information theory and and an extension of the maximum likelihood principle. In B.N Petrov and F. Csaki, editors, *2nd International Symposium on Information Theory*, pages 267–281, Budapest, 1973. Akademiai Kiado.