Applied Time Series Analysis

Vipul Bhatt

2018-08-01

Contents

Preface

These lecture notes are prepared for an upper level undergraduate course in time series econometrics. Every fall I teach a course on applied time series analysis at James Madison University. These notes borrow heavliy from the teaching material that I have developed over several years of instruction of this course.

One of my main objective is to develop a primer on time series analysis that is more accessible to undergraduate students than standard textbooks available in the market. Most of these textbooks in my opinion are densely written and assume advanced mathematical skills on the part of our students. Further, I have also struggled with their topic selection and organization. Often I end up not following the chapters in order and modify content (by adding or subtracting) to meet my students needs. Such changes causes confusion for some students and more importantly discourages optimal use of the textbook. Hence, this is an undertaking to develop a primer on time series that is accessible, follows a more logical sequencing of topics, and covers content that is most useful for undergraduate students in business and economics.

Note: These notes have been prepared by me using various sources, published and unpublished. All errors that remain are mine.

Chapter 1

Introduction to Forecasting

1.1 Time Series

A time series is a specific kind of data where observations of a variable are recorded over time. For example, the data for the U.S. GDP for the last 30 years is a time series data.

Such data shows how a variable is changing over time. Depending on the variable of interest we can have data measured at different frequencies. Some commonly used frequencies are intra-day, daily, weekly, monthly, quarterly, semi-annual and annual. Figure 1.1 below plots data for quarterly and monthly frequency.

The first panel shows data for the real gross domestic product (GDP) for the US in 2009 dollars, measured at a quarterly frequency. The second panel shows data for the retail sales in the U.S. billions of dollars, measured at monthly frequency.

Formally, we denote a time series variable by y_t, where $t = 0, 1, 2, .., T$ is the observation index. For example, at $t = 10$ we get the tenth observation of this time series, y_{10}.

1.1.1 Serial Correlation

Serial correlation is a measure of temporal dynamics of a time series. It addresses the following question: what is the effect of past realizations of a time series on the current period value? Formally,

$$\rho(s) = Cor(y_t, y_{t-s}) = \frac{Cov(y_t, y_{t-s})}{\sqrt{\sigma^2_{y_t} \times \sigma^2_{y_{t-s}}}} \tag{1.1}$$

where $Cov(y_t, y_{t-s}) = E(y_t - \mu_{y_t})(y_{t-s} - \mu_{y_{t-s}})$ and $\sigma^2_{y_t} = E(y_t - \mu_{y_t})^2$

Here, $\rho(s)$ is the serial correlation of order s. Note that often we use historical data to forecast. If there is no serial correlation, then past can offer no guidance for the present and future. In that sense, presence of serial correlation of some order is the first condition for being able to forecast a time series using its historical realizations.

1.1.2 White Noise Process

A time series is a *white noise* process is it has zero mean, constant and finite variance, and is seriall uncorrelated. Formally, y_t is a white noise process if:

1. $E(y_t) = 0$

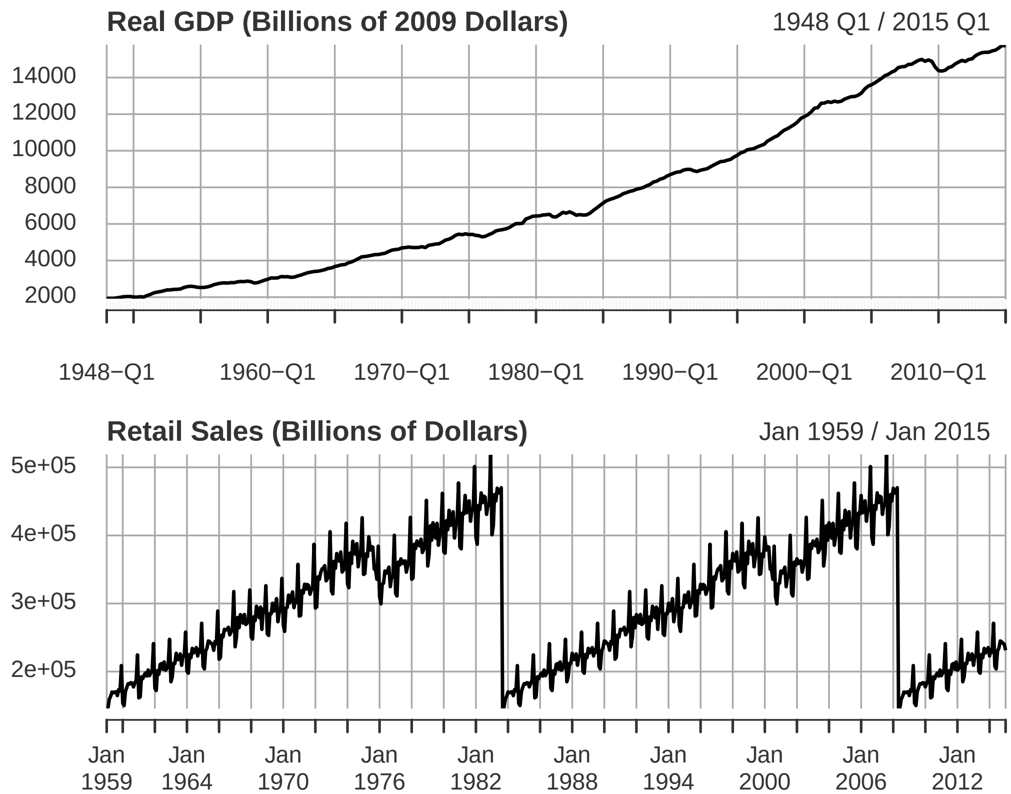

Figure 1.1: Time Series at quarterly and monthly frequency

2. $Var(y_t) = \sigma_y^2$
3. $Cov(y_t, y_{t-s}) = 0 \forall s \neq t$

We can compress the above definition as: $y_t \sim WN(0, \sigma_y^2)$. Often we assume that the unexplained part of a time series follows a white noise process. By definition we cannot forecast a white noise process. An important diagnostics of model adequacy is to test whether the estimated residuals are white noise (more on this later).

1.2 Important Elements of Forecasting

Definition 1.1 (Forecast)**.**

A *forecast* is an *informed* guess about the unknow future value of a time series of interest. For example, what is the stock price of Facebook next Monday?

There are three possible types of forecasts:

1. *Density Forecast*: we forecast the entire probability distribution of the possible future value of the time series of interest. Hence,

$$F(a) = P[y_{t+1} \leq a] \tag{1.2}$$

give us the probability that the 1-period ahead future value of y_{t+1} will be less than or equal to a. For example, the future real GDP growth could be normally distributed with a mean of 1.3% and a standard deviation of 1.83%. Figure 1.2 below plots the density forecast for real GDP growth.

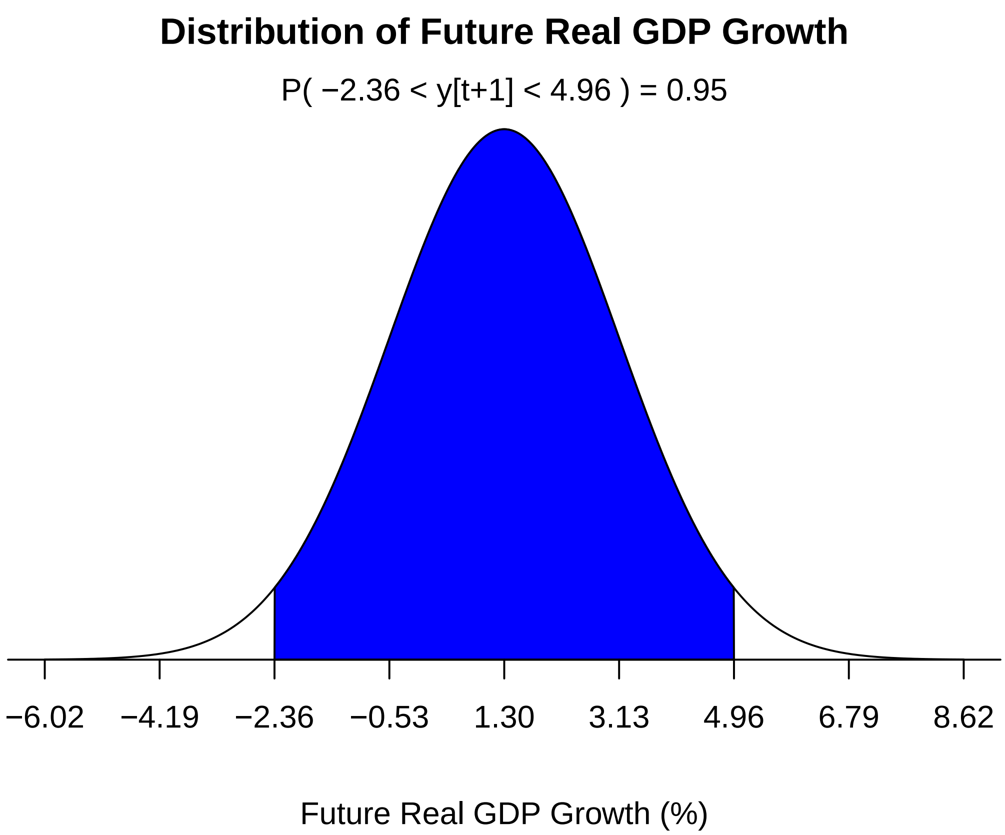

Figure 1.2: Density Forecast for Future Real GDP Growth

2. *Point Forecast*: our forecast at each horizon is a single number. Often we use the expected value or mean as the point forecast. For example, the point forecast for the 1-period ahead real GDP growth can be the mean of the probability distribution of the future real GDP growth:

$$f_{t,1} = 1.3 \tag{1.3}$$

3. *Interval Forecast*: our forecast at each horizon is a range which is obtained by adding *margin of errors* to the point forecast. With some probablity we expect our future value to fall withing this range. For example, the 95% interval forecast for the next period real GDP growth is (-2.36%,4.96%). Hence, with 95% confidence we expect next period GDP to fall between -2.36% and 4.96%.

Definition 1.2 (Forecast Horizon)**.**

Forecast Horizon is the number of periods into the future for which we forecast a time series. We will denote it by h. Hence, for $h = 1$, we are looking at 1-period ahead forecast, for $h = 2$ we are looking at 2-period ahead forecast and so on.

Formally, for a given time series y_t, the h-period ahead unknow value is denoted by y_{t+h}. The forecast of this value is denoted $f_{t,h}$.

Definition 1.3 (Forecast Error)**.**

A *forecast error* is the difference between the realization of the future value and the previously made forecast. Formally, the h-period ahead forecast error is given by:

$$e_{t,h} = y_{t+h} - f_{t,h} \tag{1.4}$$

Hence, for every horizon, we will have a forecast and a corresponding forecast error. These errors can be negative (indicating over prediction) or positive (indicating under prediction).

Definition 1.4 (Information Set)**.**

Forecasts are based on *information* available at the time of making the forecast. *Information Set* contains all the relevant information about the time series we would like to forecast. We denote the set of information available at time T by Ω_T. There are two types of information sets:

1. Univariate Information set: Only includes hisorical data on the time series of interest:

$$\Omega_T = \{y_T, y_{T-1}, y_{T-2}, ..., y_1\} \tag{1.5}$$

2. Multivaiate Information set: Includes hisorical data on the time series of interest as well as any other variable(s) of interest. For example, suppose we have one more variable x that is relevant for forecasting y. Then:

$$\Omega_T = \{y_T, x_T, y_{T-1}, x_{T-1}, y_{T-2}, x_{T-2}....., y_1, x_1\} \tag{1.6}$$

1.3 Loss Function and Optimal Forecast

Think of a forecast as a solution to an *optimization* problem. When forecasts are wrong, the person making the forecast will suffer some *loss*. This loss will be a function of the magnitude as well as the sign of the *forecast error*. Hence, we can think of an *optimal forecast* as a solution to a minimization problem where the forecaster is minimizing the loss from the forecast error.

Definition 1.5 (Loss Function)**.**

A *loss* function is a mapping between forecast errors and their associated losses. Formally, we denote the h-period ahead loss function by $L(e_{t,h})$. For a function to be used as a loss function, three properties must be satisfied:

1. $L(0) = 0$
2. $\frac{dL}{de} > 0$
3. $L(e)$ is a continuous function.

Two types of loss functions are:

- Symmetric Loss Function: both positive and negative forecast errors lead to same loss. A commonly used loss function is *quadratic loss function* given by:

$$L(e_{t,h}) = e_{t,h}^2 = (y_{t+h} - ft, h)^2 \tag{1.7}$$

- Asymmetric Loss Function: loss depends on the sign of the forecast error. For example, it could be that positive errors produce greater loss when compared to negative errors. See the function below that assumes a higher value for positive errors:

$$L(e_{t,h}) = e_{t,h}^2 + e_{t,h} \tag{1.8}$$

Once we have chosen our loss function, the optimal forecast can be obtained by minimizing the expected loss function.

Definition 1.6 (Optimal Forecast)**.**

An *optimal forecast* minimizes the expected loss from the forecast, given the information available at the time. Mathematically, we denote it by $f_{t,h}^*$ and it solves the following minimization problem:

$$min_{f_{t,h}} E(L(e_{t,h})|\Omega_t) \tag{1.9}$$

In theory we can assume any functional form for the loss function and that will lead to a different *optimal forecast*. An important result that follows from a specific functional form is stated as Theorem 1.1.

Theorem 1.1. *If the loss function is quadtratic then the optimal forecast is the conditional mean of the time series of interest. Formally, if $L(e_{t,h}) = e_{t,h}^2$ then,*

$$f_{t,h}^* = E(y_{t+h}|\Omega_t) \tag{1.10}$$

Note that $E(e_{t,h}^2)$ is known as *mean squared errors (MSE)*. Hence, the expected loss from a quadratic loss function is the same as the MSE. In this course, we assume that the forecaster faces a quadratic loss function and hence based on Theorem 1.1, we will learn different models for estimating the conditional mean of the future value of the time series of interest, i.e., $E(y_{t+h}|\Omega_t)$.

Chapter 2

Regression-based Forecasting

One way to compute the conditional expectation is the linear regression model. Here, our information set contains data on all relevant explantory variacbles available at the time of forecast, i.e,

$$\Omega_t = X_{1t}, X_{2t}, ...X_{Kt} \tag{2.1}$$

Hence, we get the following equality:

$$E(y_t|\Omega_t) = E(y_{t+h}|X_{1t}, X_{2t}, X_{3t}, ..., X_{Kt}) \tag{2.2}$$

The right hand side of the above equation is the multiple regression model of the form:

$$y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + .. + \beta_K X_{Kt} + \epsilon_t \tag{2.3}$$

We can easily estimate the above model using Ordinary Least Squares (OLS) and compute the *predicted value* of y:

$$\widehat{y_t} = \widehat{\beta_0} + \widehat{\beta_1} X_{1t} + \widehat{\beta_2} X_{2t} + ... + \widehat{\beta_k} X_{Kt} \tag{2.4}$$

The above equation can be used to compute the optimal forecast. Suppose, we are interested in computed the h period ahead forecast for y. Then, using the above equation we get:

$$\widehat{y}_{t+h} = \widehat{\beta_0} + \widehat{\beta_1} X_{1t+h} + \widehat{\beta_2} X_{2t+h} + ... + \widehat{\beta_k} X_{Kt+h} \tag{2.5}$$

2.1 Scenario Analysis and Conditional Forecasts

One way to use a regression model to produce forecasts is called *scenario analysis* where we produce a different forecast for the dependent under each possible scenario about the future values of the independent variables. For example, what will be the forecast for inflation if the Federal Reserve Bank raises the interest rate? Would our forecast differ depending on the size of the increase in the interest rate?

2.2 Unconditional Forecasts

An alternative is to separately forecast each independent variable and then compute the forecast for the dependent variable. Yet another alternative is to use lagged variables as independent variables. Depending on the number of lags, we can forecast that much ahead into future (see Distributed Lag Section for details).

2.3 Some practical issues

1. To forecast the dependent variable we first need to compute a forecast for the independent variable. Errors in this step induce errors later.
2. *Spurious regression*: It is quite possible to find a strong linear relationship between two completely unrelated variables over time if they share a common time trend.
3. *Model Uncertainty*: We do not know the true functional form for the regression model and hence our estimated model is only a proxy for the true model.
4. *Parameter Uncertainty*: This kind of forecast uses regression coefficients that are computed using a fixed sample. Over time with new data, there will be changes in these coefficients.

2.4 Distributed Lag Regression Models

Consider the following simple regression model:

$$y_t = \beta_0 + \beta_1 x_t + \epsilon_t \tag{2.6}$$

Here, if want to forecast y_{t+1} then we must either consider different scenarios for x_{t+1} or indpendently forecast x_{t+1} first, and then use it to compute forecast for y_{t+1}. An alternative is to estimate the following lagged regression model:

$$y_t = \beta_0 + \beta_1 x_{t-1} + \epsilon_t \tag{2.7}$$

Note that by estimating the above model we get the following predicted value equation for $t+1$:

$$\widehat{y_{t+1}} = \widehat{\beta_0} + \widehat{\beta_1} x_t \tag{2.8}$$

Hence, we can easily produce 1-period ahead forecast from this model. In order to produce forecast farther into future we would need to add more lags of the independent variable to the model. A generalized model of this kind is called *distributed lag model* and is given by:

$$y_t = \beta_0 + \sum_{s=1}^{p} \beta_s x_{t-s} + \epsilon_t \tag{2.9}$$

The number of lags to include can be determined using some kind of goodness of fit measure.

2.5 Model Selection Criterion

Most often we compare models that have different number of independent variables. Here, we must account for the tradeoff between goodness of fit and degrees of freedom. Increasing the number of independent variables will:

1. lower the MSE and hence leads to better fit.
2. lowers the degrees of freedom

Two commonly used measures based on MSE incorporate this tradeoff:

1. Akaike Information Criterion (AIC):

$$AIC = MSE \times e^{\frac{2k}{T}}$$

where k is the number of estimated parameteris, T is the sample size. Then, K/T is the number of parameters estimated per observation and $e^{\frac{2k}{T}}$ is the *penalty factor* imposed on adding more variables to the model. As we increase k, this penalty factor will increase exponentially for a giveb value of T.

2. Bayesian Information Criterion (BIC):

$$BIC = MSE \times T^{\frac{k}{T}}$$

Lower values of either AIC or BIC indicates greater accuracy. So we select a model with lower value of either of these two criteria. Note that the penalty imposed by BIC is harsher and hence it will typically select a more parsiomonius model.

Chapter 3

Components of a Time Series

A given time series can have four possible components:

1. Trend: denoted by B_t captures the long run behavior of the time series of interest.
2. Season: denoted by S_t are *periodic* fluctuations over *seasons*. The period of the season is fixed and known. For example, rise in non-durable sales during Christmas.
3. Cycle: denoted by C_t are *non-periodic* are flcutuations in that they occur regularly but over periods that are not fixed in duration.
4. Irregular: denoted by ϵ_t are random fluctuations, typically modeled as a white noise process.

3.1 Decomposing a time series

We can decompose any given time series into its components. There are two ways to accomplish this:

1. Additive Decomposition: Here it is assumed that all four components are added to obtain the underlying timer series:
$$y_t = B_t + S_t + C_t + \epsilon_t \tag{3.1}$$
2. Multiplicative Decomposition: Here it is assumed that all four components are multiplied to obtain the underlying timer series:
$$y_t = B_t \times S_t \times C_t \times \epsilon_t \tag{3.2}$$

Note that using properties of logarithms, multiplicative decomposition is the same as additive decomposition in log terms:

$$log(y_t) = log(B_t) + log(S_t) + log(C_t) + log(\epsilon_t) \tag{3.3}$$

Most statistical softwares can implement these decompositions using data on a time series variable as input. Typically they combine cyclical component with irregular component and provide a three-way decomposition. In Figure 3.1 I use R to decompose real GDP for the US into its components.

3.2 Uses of Decomposition of a time series

The usefulness of decomposing a time series depends on our objective.

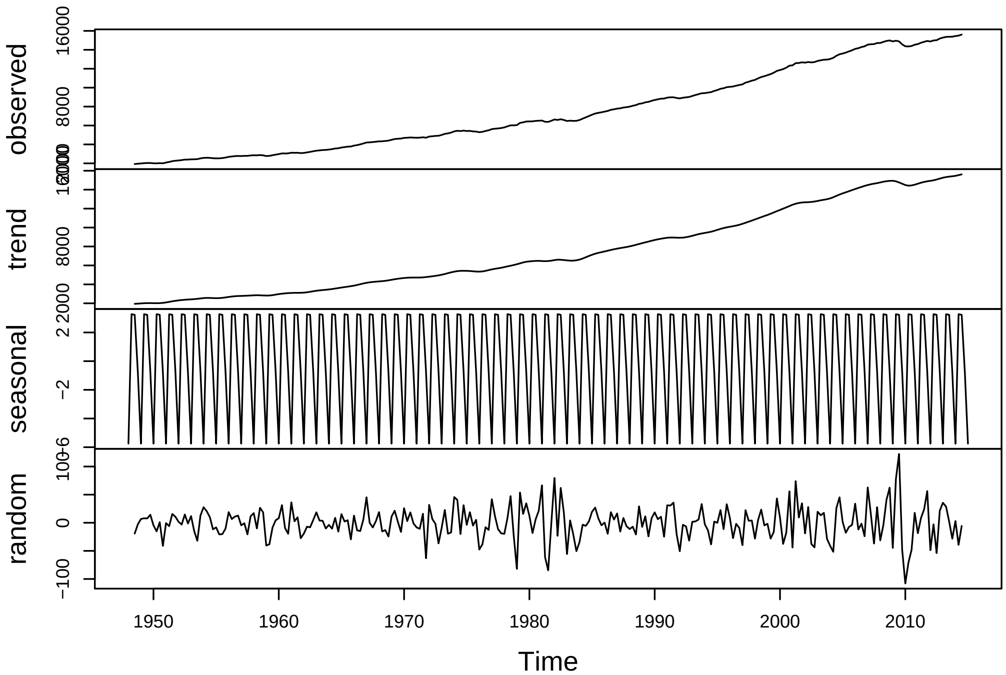

Figure 3.1: Additive Decomposition of Real GDP

1. It may be of interest to study each component separtaely or to simply improve our understanding of the temporal dynammics of a time series of interest. Decompositing it into different components is the first step towards achieving that goal.

2. We can also use the decomposition to filter out components that we are not interested in studying. If for example we are only interested in modeling the cyclical component of the time series, then we can assume some kind decomposition, additive or multiplicative, and filter out the trend and seasonal component. For example, assuming additive decomposition, the filtered time series is given by:

$$Filtered\ y_t = y_t - B_t - C_t \tag{3.4}$$

We can then proceed to model the cyclical component using the filtered data.

Chapter 4

Smoothing Methods

One way to approach forecasting is to *average* out the fluctuations in the underlying time series to produce a *smoothed* data which can be extrapolated to produce forecasts. These smoothing methiods are essentially *model-free* and may not even produce *optimal forecasts*. Depending on the method used one can accomodate seasonal as well as trend components of the underlying time series.

4.1 Moving Average Method

We compute an average of most recent data values for the time series and use it as a forecast for the next period.

An important parameter is the *window* over which we take the average. Let us denote this window by m, then:

$$y^f_{T+1} = \frac{\sum_{i=t-m+1}^{t} y_i}{m} \tag{4.1}$$

A larger value of m produces greater smoothing and most softwares have a default value of this parameter which can be changed if needed.

4.2 Simple Exponential Smoothing

In the moving average method, all obseravtions receieved same weight. However, it is reasonable to argue that more recent observations may have a greater influence than those in the remote past. In this methiod, the weight attached to past observations exponentially decay over time. Here is the algorthim for computing the smoothed data and its forecast:

1. Initialize at t=1:

$$y^s_1 = y_1$$

2. Update:

$$y^s_t = \alpha y_t + (1-\alpha) y^s_{t-1} \quad for\ t = 2, 3, ...T$$

3: h-period ahead forecast:

$$f_{T,h} = y^s_T$$

Here the h-period ahead forecast is:

Exercise: Can you show that y_t^s is a is the weighted moving average of all past observations? Use backward substitution method.

Here $\alpha \in (0, 1)$ is the smoothing parameter, with smaller value indicating greater smoothing.

4.3 Holt-Winters Smoothing

We add trend component to the simple exponential smoothing. In step 2 the equation we use to update the smoothed data is given by:

$$y_t^s = \alpha y_t + (1-\alpha)(y_{t-1}^s + B_{t-1})B_t = \beta(y_t^s - y_{t-1}^s) + (1-\beta)B_{t-1} \tag{4.2}$$

We now have an additional parameter β that is the trend parameter. Here the h-period ahead forecast is:

$$f_{T,h} = y_T^s + h \times B_T \tag{4.3}$$

4.4 Holt-Winters Smoothing with Seasonality

We now add seasonal component along with trend. Assuming multiplicative seasonality with period n:

$$y_t^s = \alpha \frac{y_t}{S_{t-n}} + (1-\alpha)(y_{t-1}^s + B_{t-1})B_t = \gamma(y_t^s - y_{t-1}^s) + (1-\gamma)B_{t-1}S_t = \beta \frac{y_t}{y_t^s} + (1-\beta)S_{t-n} \tag{4.4}$$

The h-period ahead forecast is given by:

$$f_{T,h} = (y_T^s + h \times B_T) \times S_{T+h-n} \tag{4.5}$$

4.5 Application

We use R to implement a 12-period ahead forecast for retail sales for the U.S. The data is at monthly frequency from 1959 through 2015. We use natural logs of the data and compute smoothed series using each of the three methods. The resulting forecasts are ploted in Figure 4.1.

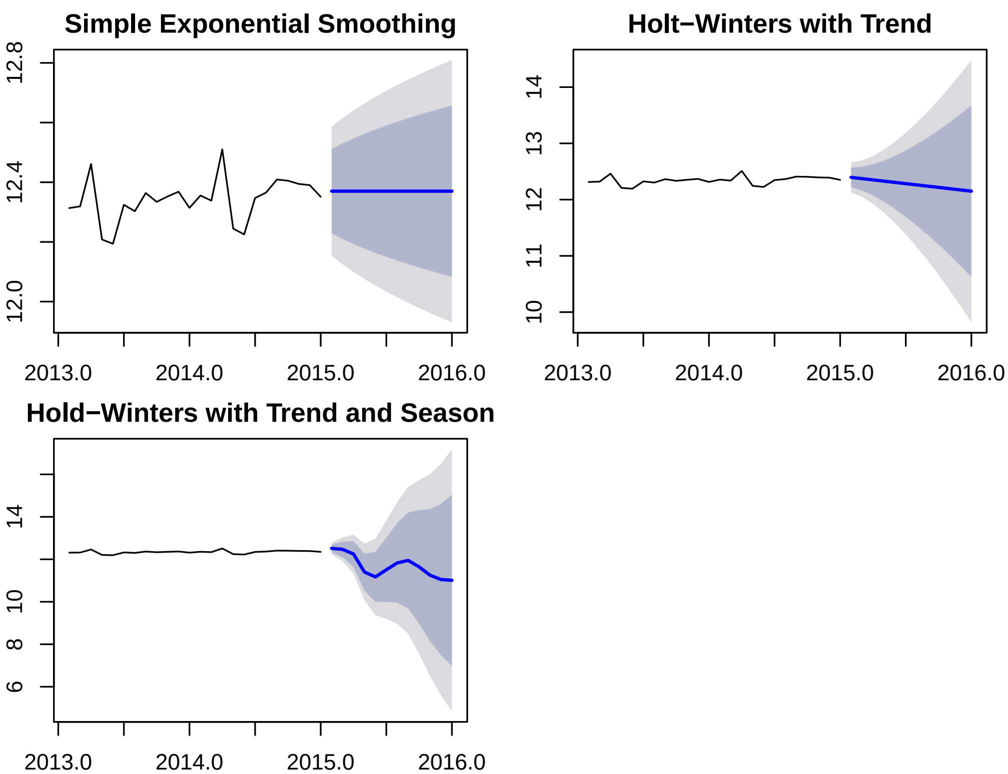

Figure 4.1: Forecast of Retail Sales: Three Smoothing Methods

Chapter 5

Modeling Trend and Seasonal Components

5.1 Trend Estimation

An important component of a time series is *trend* that captures the long run evolution of the variable of interest. There are two types of trends:

1. Deterministic Trend: the underlying trend component is a *known* function of time with *unknown* parameters.
2. Stochastic Trend: the trend component is random.

In this note we will focus on estimating and forecasting deterministic trend models. We will come back to stochastic trend later when we talk about stationarity property of a time series.

5.1.1 Parametrizing a deterministic trend

Whether or not there is deterministic trend in the data can be typically gleaned by simply plotting the time series over time. For example, Figure @ref(fig: figurefive) below plots real GDP for the US at quartely frequency. We can observe a positive time trend with real GDP increasing with time. In this section we will learn to *fit* a function that captures this relationship accurately.

Note: The variable time is denoted by t and it is artificially created to take value of 1 for the first period, 2 for the second period and so on.

There are two commonly used functional forms for capturing a deterministic trend:

1. Polynomial Trend: We fit a polynomial of appropriate order to capture the time trend. For example, A. Linear trend:

$$y_t = \beta_0 + \beta_1 t + \epsilon_t \tag{5.1}$$

B. Quadratic trend:

$$y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \epsilon_t \tag{5.2}$$

In general, we can fit a polynomial of order q:

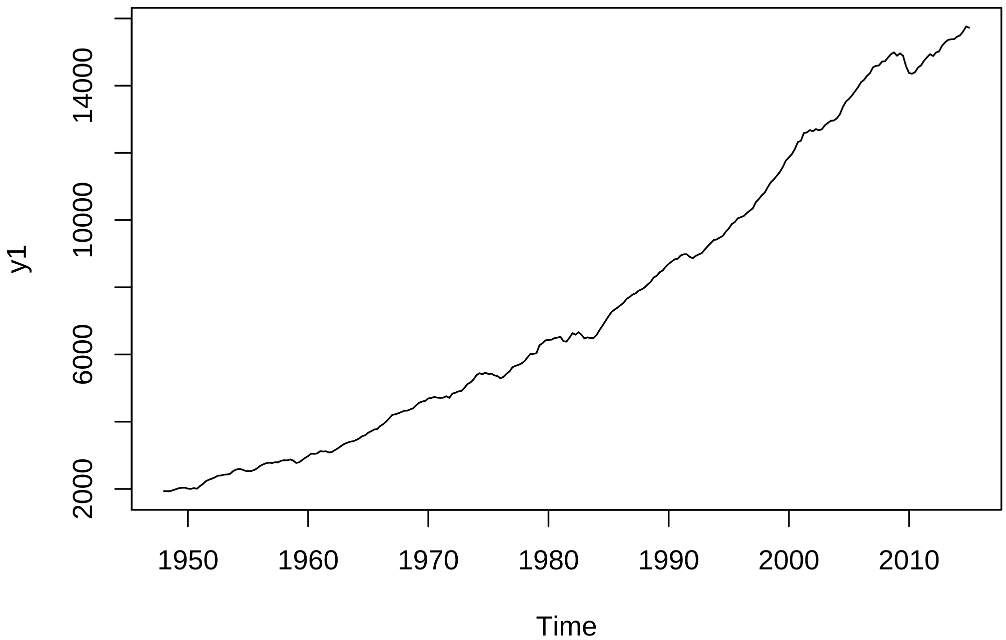

Figure 5.1: Plot of Real GDP

$$y_t = \beta_0 + \sum_{i=1}^{q} \beta_i t^i + \epsilon_t \tag{5.3}$$

We can estimate this model using the OLS. One of the key component here is to determine the *right* order of the polynomial. We can begin with a large enough number for q and then select the approprirate order using AIC or BIC criterion.

2. Exponential or log-linear trend: In some cases we may want to use an exponential trend or equivalently a log-linear trend.

$$y_t = e^(\beta_0 + \beta_1 t + \epsilon_t) equivalently log(y_t) = \beta_0 + \beta_1 t + \epsilon_t \tag{5.4}$$

Again we can estimate the above model using OLS.

5.1.2 Uses of the Deterministic Trend Model

Once we have finalized our deterministic trend model i.e., either a polynomial of a specific order or log-liner trend, we can use the estmated model for the following two purposes:

1. Detrending our data: Suppose we would like to eliminate trend from our data. The residual from our final trend model is the *detrended* time series.

2. Forecasting: We can also forecast our time series based on the estimated trend. For example, suppose our final model is a quadratic trend. The predicted value is given by:

$$\widehat{y_t} = \widehat{\beta_0} + \widehat{\beta_1} t + \widehat{\beta_2} t^2 \tag{5.5}$$

Then, the 1-period ahead forecast for y_{t+1} can be obtained by solving:

$$\widehat{y_{t+1}} = \widehat{\beta_0} + \widehat{\beta_1}(t+1) + \widehat{\beta_2}(t+1)^2 \tag{5.6}$$

Table 5.1: Optimal Order of the Polynomial

order	AIC	BIC
1	4399.225	4410.009
2	3877.692	3892.071
3	3868.441	3886.415
4	3763.525	3785.094

Table 5.2: Regression Results

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	7520.008	15.898	473.021	0
poly(t, 4)1	69434.197	260.744	266.293	0
poly(t, 4)2	12928.805	260.744	49.584	0
poly(t, 4)3	-1068.061	260.744	-4.096	0
poly(t, 4)4	-2959.656	260.744	-11.351	0

5.1.3 Application: Estimating a polynomial trend for U.S. Real GDP

We will now fit a polynomial trend to the US real GDP data that was presented in Figure 5.1. We first estimate polynomials of different orders and select the optimal order determined by the lowest possible AIC/BIC. Table 5.1. shows these statistics for upto 4th order polynomial. We find that the lowest value occur at $q = 4$.

Hence, our final trend model is:

$$y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \beta_4 t^4 + \epsilon_t \tag{5.7}$$

The estimated trend model is presented in Table 5.2.

Using the estimated model, we can compute the detrended data as the residual and also forecast y_t. Figure 5.2 below plots the detrened real GDP obtained as a residual from our trend model.

Table 5.3 shows the forecast of real GDP for next 8 quarters along with the 95% confidence bands.

Table 5.3: Forecast for 8 Quarters

h	Forecast	95% Lower Bound	95% Upper Bound
1	15924.30	15764.22	16084.36
2	15968.97	15801.63	16136.31
3	16012.42	15837.52	16187.31
4	16054.60	15871.86	16237.34
5	16095.50	15904.62	16286.38
6	16135.09	15935.78	16334.40
7	16173.35	15965.32	16381.38
8	16210.24	15993.19	16427.29

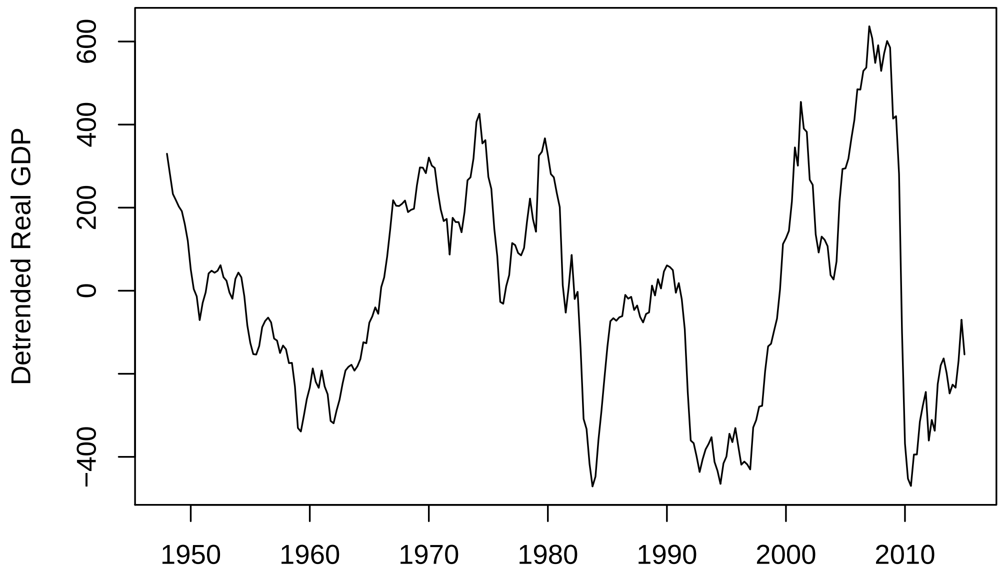

Figure 5.2: Plot of Detrended Real GDP

5.2 Seasonal Model

Chapter 6

Autoregressive Moving Average (ARIMA)

6.1 Covariance Stationary Time Series

Definition 6.1 (Covariance Stationary Time Series)**.**

A time series $\{y_t\}$ is said to be a *covariance stationary process* if:

1. $E(y_t) = \mu_x \quad \forall \quad t$
2. $Var(y_t) = \sigma_x^2 \quad \forall \quad t$
3. $Cov(y_t, y_{t-s}) = \gamma(s) \quad \forall \quad s \neq t$

6.2 Correlation over time

In general, for a time series, $\{y_t\}$,

$$Cor(y_t, y_{t-s}) = \frac{Cov(y_t, y_{t-s})}{\sqrt{\sigma_{y_t}^2 \times \sigma_{y_{t-s}}^2}} \tag{6.1}$$

where $Cov(y_t, y_{t-s}) = E(y_t - \mu_{y_t})(y_{t-s} - \mu_{y_{t-s}})$ and $\sigma_{y_t}^2 = E(y_t - \mu_{y_t})^2$

Definition 6.2 (Auto Correlation Function (ACF))**.**

An *ACF* plots the correlation of a time series with its own past values over time. For a stationary time series, using the three conditions we get:

$$ACF(s) \ or \ \rho(s) = \frac{\gamma(s)}{\gamma(0)} \tag{6.2}$$

Definition 6.3 (Partial Auto Correlation Function (PACF))**.**

The *partial autocorrelation function (PACF)* for a stationary time series y_t at lag s is the direct correlation between y_t and y_{t-s}, after filtering out the linear influence of $ y_{t-1},…,y_{t-s-1}$ on y_t.

6.3 Autoregressive (AR) Model

A *stationary*time series $\{x_t\}$ can be modeled as an AR(p) process:

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + \epsilon_t \tag{6.3}$$

6.4 Moving Average (MA) Model

A *stationary* time series $\{y_t\}$ can be modeled as an MA(q) process:

$$y_t = \theta_0 + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + ... + \theta_q \epsilon_{t-q} \tag{6.4}$$

6.5 ARMA(p, q)}

An ARMA model simply combines both AR and MA components to model the dynamics of a time series. Formula,

$$y_t = \phi_0 + \phi_1 x_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + ... + \theta_q \epsilon_{t-q} \tag{6.5}$$