

# Applied Time Series Analysis

*Vipul Bhatt*

*2018-07-20*



# Contents

<b>Preface</b>	<b>5</b>
<b>1 Introduction</b>	<b>7</b>
<b>2 Smoothing Methods and Regression-based Forecasting</b>	<b>9</b>
2.1 Smoothing Methods . . . . .	9
2.2 Regression-based Forecasting . . . . .	10



# 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 heavily 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*Note: These notes have been prepared by me using various sources, published and unpublished.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

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  $x_t$ , where  $t = 0, 1, 2, \dots, T$  is the observation index. For example, at  $t = 10$  we get the tenth observation of this time series,  $x_{10}$ .

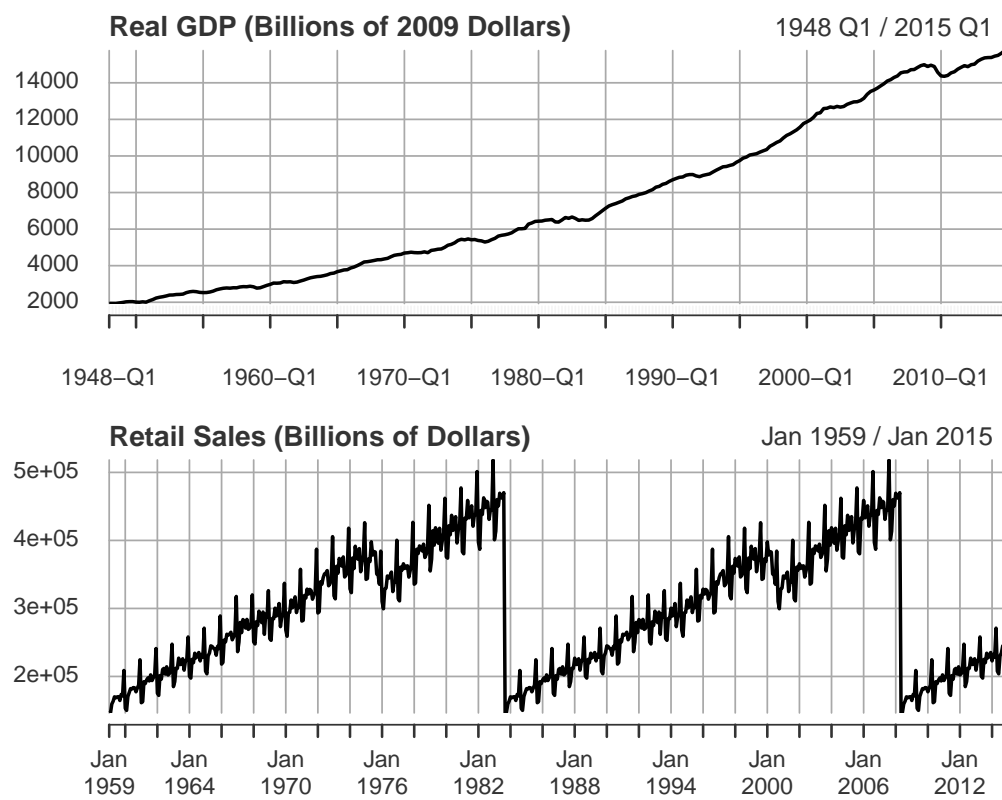


Figure 1.1: Time Series at quarterly and monthly frequency



## Chapter 2

# Smoothing Methods and Regression-based Forecasting

In this chapter we will look at simple techniques that can be used to forecast a time series variable. Here we focus on methods that do not account for various components of a time series explicitly. The list of techniques covered in this topic are:

1. Smoothing Methods
2. Regression-based Forecasting

“Naive Forecasting Methods” There are many ‘back-of-envelope’ type of forecasting methods often used by practitioners. For example:

### 2.1 Smoothing Methods

This approach attempts to *average* out the irregular component of a Time series.

#### 2.1.1 Moving Average Method

Here 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:

$$y_{T+1}^f = \frac{\sum_{i=t-m+1}^t y_i}{m} \quad (2.1)$$

As we increase  $m$ , more weight is given to more recent observations and hence we get less *smoothing*. A larger value is more desirable when there are large but infrequent fluctuations in our data. In contrast, a smaller value of  $m$  is preferred when data has sudden shifts in our data. Typically, we set  $m$  equal to the frequency at which our data is measured. For example,  $m = 4$  for quarterly data and  $m = 12$  for monthly data. One drawback of this method is that it assigns equal *weight* to each observation. It is reasonable to argue that for most economic and financial variables, the effect of past observations will be less pronounced than most recent observations.

### 2.1.2 Simple Exponential Smoothing

Here, the weight attached to past observations exponentially decay over time. The next period forecast is the exponentially weighted moving average of all previously observed values.

$$y_t^s = \alpha y_t + (1 - \alpha)y_{t-1}^s \quad (2.2)$$

Here the h-period ahead forecast is:

$$y_{T+h}^f = y_T^s \quad (2.3)$$

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

### 2.1.3 Holt's Exponential Smoothing

Adds trend component to simple exponential smoothing.

$$y_t^s = \alpha y_t + (1 - \alpha)(y_{t-1}^s + B_{t-1})B_t = \gamma(y_t^s - y_{t-1}^s) + (1 - \gamma)B_{t-1} \quad (2.4)$$

Here the h-period ahead forecast is:

$$y_{T+h}^f = y_T^s + h \times B_T \quad (2.5)$$

### 2.1.4 Winter's Exponential Smoothing

Adds seasonal component along with trend. Assuming multiplicative seasonality:

$$y_t^s = \alpha \frac{y_t}{S_{t-n}} + (1 - \alpha)(y_{t-1}^s + B_{t-1})B_t = \gamma(y_t^s - y_{t-1}^s) + (1 - \gamma)B_{t-1}S_t = \beta \frac{y_t}{y_t^s} + (1 - \beta)S_{t-n} \quad (2.6)$$

where  $n$  is the number of periods in a season.

$$y_{T+h}^f = (y_T^s + h \times B_T) \times S_{T+h-n} \quad (2.7)$$

## 2.2 Regression-based Forecasting

A linear regression model estimates the value of the dependent variable as a function of the independent variable. The predicted value of the dependent variable can be used as a “forecast” when working with the time series data. For example:

$$Y_t = \beta_0 + \beta_1 X_t + \epsilon_t \quad (2.8)$$

Then given a sample of observations over time for  $Y$  and  $X$ , we can estimate the above model using OLS and compute the predicted value of  $Y$ :

$$\hat{Y}_t = \hat{\beta}_0 + \hat{\beta}_1 X_t \quad (2.9)$$

Now suppose we are interested in computed the  $h$  period ahead forecast for  $Y$ . Then, using the above equation we get:

$$\widehat{Y}_{T+h} = \widehat{\beta}_0 + \widehat{\beta}_1 X_{T+h} \quad (2.10)$$

Hence, to forecast the dependent variable we first need to compute a forecast for the independent variable. Larger the number of independent variables in our model greater the number of forecasts we need to compute making regression-based forecasting a rather difficult approach to implement in practice.

Further, when working with time series data, we have to be careful about *spurious regression* problem. It is quite possible to find a strong linear relationship between two completely unrelated variables over time if they share a common stochastic trend.