Economic Forecasting
Lecture 1: Introduction to Forecasting

Richard G. Pierse

1 Introduction

Forecasting is an important part of life. In our everyday life we are constantly making forecasts, often even without being conscious of it. When deciding whether or not I can afford to buy something, I need to be able to forecast the state of my bank account when the bill will be due. When I leave my house in the morning, I make a forecast of the probability of rain in order to decide whether or not to carry an umbrella. If I am thinking of buying a house, I need to be able to forecast many things: my future income, future interest rates, future house prices, the future size of my family etc. before I come to a decision.

In business, all firms need to make forecasts of future sales in order to make decisions about production, inventory levels, employment etc. Public companies are obliged to publish profits forecasts for the benefit of their shareholders. Future sales and profits for a firm will be influenced, not only by the state of the firm itself, but by what is happening in the industry, in the economy as a whole and, for firms importing raw materials or exporting final goods, the state of the world economy. This means that the firm needs to have forecasts at all these levels although typically a firm will not build its own forecasts beyond the level of the firm or industry but instead will rely on publicly available forecasts or buy in forecasts from outside economic consultants.

Governments and Central Banks need to forecast the main macroeconomic variables in order to make decisions about setting tax rates and interest rates. Some of these forecasts are made available to the public. For example, the Bank of England publishes quarterly forecasts of inflation and output growth in its *Inflation Report*. The economic model that underpins these forecasts is also set out in a publicly accessible document. The UK Treasury publishes forecasts of public expenditure and while it does not publish its own macroeconomic forecasts, it does publish a summary of independent forecasters.

It can be seen from the examples considered above that forecasts are made in order to guide decisions. A good forecast will help to produce a good decision and a bad forecast can lead to a bad decision.

2 The Forecast Object

There are many different types of object that we might potentially want to forecast. In business and economics, the forecast object is typically one of three types: an *event outcome*, an *event timing* or the *values* of some economic variable. For example we might be interested in forecasting the outcome of the next decision by the *Monetary Policy Committee* (*MPC*) on whether *UK* interest rates will be raised, lowered or unchanged. This would be an event outcome forecast with three possible outcomes. Alternatively, we might be interested in forecasting exactly *when* interest rates will next be increased; this is an event timing. Finally, we might be interested in forecasting what interest rates will be in twelve months time. In this case we need to forecast the *values* of interest rates one year ahead. This third type of forecast is also called a *time series forecast*.

If we are able to forecast the time series of future interest rates far enough ahead, then clearly we can answer all of the three questions considered above. However, it may not be necessary to produce a time series forecast in order to answer event outcome or event timing questions. In practice, *time series forecasts* are the most frequently encountered type of forecast and the type that we will concentrate on for the rest of this module. The techniques for time series forecasting are well developed and time series forecasts can be made and evaluated routinely. However, it should be noted that alternative forecasting procedures (perhaps non-quantitative) might be more appropriate if we are primarily interested in *event outcome* or *event timing* forecasts.

3 Evaluating Forecasts: the Loss Function

Forecasts are made in order to to help take decisions. The consequence of a bad forecast is that a bad decision may be taken and that leads to a loss. If I decide to leave my umbrella at home and it rains, then I get wet. If I take my umbrella and it doesn't rain, then there is a loss associated with my carrying about an unnecessary implement. Most people would probably judge that the former loss is greater than the latter, in which case the loss is *asymmetric* with respect to the two wrong decisions.

Consider the stylised problem faced by a firm which has to decide how much inventory to hold for next period. If demand next period is high, then the firm would want to have a high inventory level or it will be unable to satisfy demand. If demand next period is low, then the firm would like to hold low inventory since holding inventory costs money. Since the firm does not know what demand will be next period, it needs to make a forecast of demand and make its decision on inventory holding on the basis of this forecast. If it makes a wrong decision, then a loss will be incurred. The decisions and loss outcomes can be represented in a table:

	Demand High	**Demand Low**
High inventory	0	£1000
Low inventory	£1000	0

The diagonal cells in the table represent the case where the firm makes the right decision and the loss is zero. The off-diagonal elements represent the two cases where the firm makes a wrong decision and incurs a loss. In this case the loss is the same in each case. This is a *symmetric* loss structure. In many decision environments, a symmetric loss structure may closely approximate the true losses of the forecaster. In other cases, a symmetric loss structure may not be realistic. In this case, it might be thought that the loss from losing out on demand by too low a level of inventory might considerably outweigh the loss from holding too much inventory in a period of low demand. This can be represented in the alternative loss table which displays an *asymmetric* loss structure:

	Demand High	**Demand Low**
High inventory	0	£1000
Low inventory	£3000	0

In the stylised example just considered, there were only two possible forecasts (high demand or low demand) and only two decisions (high inventory or low inventory). In the more realistic case of *time series forecasting*, both forecast and outcome can assume a continuous range of values. Suppose that we are interested in forecasting a variable, y, in time period t.Denoting the forecast of y_t as $\widehat{y}_t$, the difference between the forecast and the outcome is given by

$$e_t = y_t - \widehat{y}_t$$

where e_t is the corresponding *forecast error.*

The loss associated with the forecast is a function of the forecast error and we write it as

$$L(e_t).$$

A loss function should satisfy the properties that: (i) $L(0) = 0$, so that no loss is incurred when the forecast is perfect, (ii) $L(e_t)$ is *increasing* in $|e_t|$ so that the loss increases with the absolute size of the forecast error and (iii) $L(e_t)$ is continuous so that nearly identical forecast errors produce nearly identical losses.

Two important loss functions are the *quadratic loss function*

$$L(e_t) = e_t^2$$

and the *absolute loss function*

$$L(e_t) = |e_t|.$$

The quadratic loss function associates a much higher loss to large forecast errors than small errors whereas for the absolute loss function the loss is linear in the forecast error. Both loss functions are symmetric so that positive and negative forecast errors are weighted equally. In some circumstances as we have seen, asymmetric loss functions may be more realistic but in practice the quadratic loss function is the most widely used.

Given a loss function, an *optimal forecast* be defined as a forecast that minimises the expected loss. This will be an important concept in subsequent lectures.

4 Forecast Types

A forecast can be presented in any one of several different ways: as a best guess, as a range within which the value is expected to fall with a certain probability or as a whole probability distribution for the future value. In the first case the forecast is called a *point forecast*, in the second case an *interval forecast* and in the third case a *density forecast*.

4.1 Point forecasts

A *point forecast* is a single number e.g. output growth next year is expected to be 3%, sales for next month are expected to be up by 5%. This is a very simple way of presenting a forecast and is handy in many ways: for example a point forecast of one variable can be used to help forecast a second variable. However, a point forecast abstracts from the uncertainty associated with the forecast and gives no way to measure the size of that uncertainty. Suppose that we are shown two different forecasts for next year's output growth, 2.5% and 3.5%. Are these two forecasts significantly different from each other? Without more information about the uncertainty of each forecast, it is impossible to know.

4.2 Interval forecasts

An *interval forecast* is a range of values within which the realised future value is expected to fall with some certain probability. The probability may not always be specified explicitly. An example is: 'with 90% probability, inflation for the next quarter will be between 2% and 3%'. The length of the interval gives a measure of the uncertainty of the forecast: the wider the interval the less certain is the forecast. In this respect an interval forecast contains more information than a point forecast. A point forecast can be derived from an interval forecast if we assume that the underlying probability distribution is symmetric. In this case a point forecast would just be given by the centre point of the interval. Sometimes a symmetric interval forecast may be expressed as a central forecast plus or minus a figure e.g. 3.0 ± 1.5 representing the interval $[1.5, 4.5]$. However, interval forecasts are not necessarily symmetric.

4.3 Density forecasts

A density forecast gives the entire density or probability distribution of the future value of the series. This provides complete information about the uncertainty associated with the forecast. A good example of a density forecast is the inflation forecast published by the Bank of England in their quarterly *Inflation Report*. This is presented in the form of a *fan chart* (known informally as the 'rivers of blood' chart). The chart (Figure 1) represents different deciles from the forecast

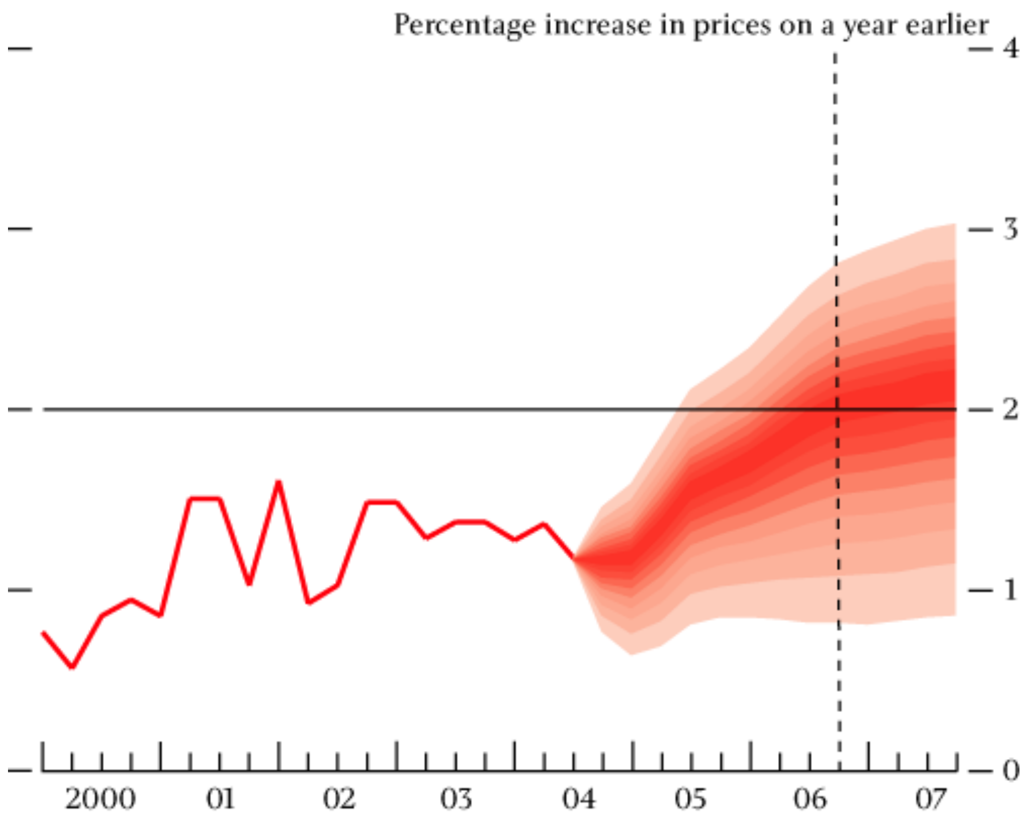

Figure 1: Inflation fan chart: *Bank of England Inflation Report November 2004*

distribution of inflation, each shown in a different shade of red. Interval forecasts can be derived from the fan chart for each of the percentiles shown. One feature of the Bank of England forecast distribution is that it is *not symmetric* so that intervals corresponding to a particular probability are not equally distributed around the central point forecast that represents the 'most likely' outcome. This represents an explicit view that the 'upside' and 'downside' risks around the central forecast estimate are not equal. (Other fan charts such as those reported by the National Insitute of Economic and Social Research do not have this feature). One other obvious feature is the fanning out of the distribution over time, which illustrates the increased uncertainty of inflation forecasts as we look further ahead.

5 The Forecast Horizon

The forecast horizon is the number of periods between when the forecast is made and the period that is being forecasted. We use the notation

$$\widehat{y}_{t+h,t-s}$$

to denote a forecast of y_{t+h} made in period $t-s$. The forecast horizon here is $s+h$ and the forecast is known as an $s+h$*-step ahead forecast*. Clearly the number of steps ahead will depend on the data frequency so that a forecast of sales a year hence will be a one-step forecast if using annual data but a twelve-step forecast if using monthly data. Generally, the longer is the forecast horizon, the more uncertain will be the forecast.

The choice of how far ahead to forecast is an important one. Since forecasts are made primarily to influence decisions, the forecast must be made far enough ahead that the appropriate decision can be implemented in time. For example, if it takes two years for a firm to build a new factory in order to increase its production, then the forecasts that influence its production decision need to be made at least two years ahead. On the other hand, if the forecasts are made too far ahead, then they may be unreliable.

The forecast horizon may influence the type of model needed to produce an appropriate forecast. For example, a reasonable forecast of the temperature in Guildford tomorrow can probably be made just on the basis of local weather fronts today. On the other hand, to make a reasonable forecast of the temperature in Guildford in fifty years time we need to consider other factors such as global warming. Similarly, in economics we have both short-run models and medium-run or long-run models. In the short-run models, some factors that only change slowly over time such as the labour force or the capital stock, are treated as constant. If the forecast horizon is short, then a short-run model may provide a good enough forecast. However, if the forecast horizon is longer, then it will be more important to use a medium-run or long-run model.

6 The Information Set

The quality of a forecast will depend on the quality and the quantity of information available when the forecast is made. The *information set* is the set of all information used to produce a forecast and is denoted as $\mathbf{\Omega}_t$. Note that the information set is indexed by t to indicate that this is the set of information available at time t. Suppose that we are interested in forecasting the variable y_{t+h} in period t. One source of available information is *past* (and possibly *present* values) of the variable y itself. The *univariate information set* is the set of all past values of y:

$$\mathbf{\Omega}_t = \{y_t, y_{t-1}, y_{t-2}, \cdots\}.$$

Note that here it is assumed that y_t is part of the information set available to forecasters in period t. This is equivalent to assuming that the forecast is made at the *end* of the period after the current value becomes available. Alternatively, we could assume that the forecast is made at the *start* of the period in which case y_t would not be part of $\mathbf{\Omega}_t$. In reality, it takes time to collect data so that the latest observation in the information set $\mathbf{\Omega}_t$ may be several periods earlier.

In addition to past values of y, a forecast may use past (and present) values of other variables such as x. The *multivariate information set*

$$\mathbf{\Omega}_t^* = \{y_t, x_t, y_{t-1}, x_{t-1}, y_{t-2}, x_{t-1}, \cdots\}$$

is *larger* than the univariate information set $\mathbf{\Omega}_t$ since it includes additional information.

When comparing different forecasts, we must take account of the fact that they may be based on different information sets. A forecast made at a later date is based on a larger information set so that, for example, the forecast $\widehat{y}_{t+2,t+1}$ is based on more information than $\widehat{y}_{t+2,t}$. Increasing the information set should not make a forecast worse but it might not improve a forecast. Choosing the appropriate information set therefore, for example univariate or multivariate, is a crucial decision in forecasting. We will look at both univariate and multivariate forecasting methods in the course of this module.

7 The Parsimony Principle

The *parsimony principle* states that, *other things being equal*, a simple model is better than a complex model. This principle is also known as *Occam's razor* after the medieval philosopher William of Occam (*ca.* 1285–1349), born in Ockham in Surrey, who first propounded it in the latin form '*pluralitas non est ponenda sine necesitate*' or 'plurality should not be conceded unnecessarily'. Another popular version of the same idea is the *KISS principle*, standing for '*Keep It Simple, Stupid*'.

In forecasting, professional experience over many decades has suggested that simple, parsimonious models can out-perform more complicated models. In particular, univariate models often seem to do better for short term forecasting than multivariate models, despite being based on a smaller information set. One possible reason for this is that, when forecasting more than one period ahead, a multivariate model needs to make forecasts for all its model variables. Errors in forecasting one variable can feed into forecasts of other variables in the system.

One other reason why simple models may forecast better is the issue of '*over-fitting*'. Forecasting models normally involve unknown parameters that need to be estimated. Increasing the complexity of a model by adding extra parameters, cannot reduce the fit of the model *within-sample*, that is over the period used for estimation, and will generally improve it. However, a model that has been *over-fitted* will generally forecast badly *out-of-sample* because the estimated parameters have been tailored to pick up *ideosyncracies* of the historical estimation period that are not relevant to future observations outside the sample. Luckily, statistical criteria are available that can guide model selection within classes of models and help us avoid over-fitting models.

8 Some Naïve Forecasting Models

Let us take a look at some naïve forecasting models. These models are based on the projection of very simple rules that do not involve any unknown parameters to be estimated. By the criterion of parsimony these models score well and when the forecast horizon is very short, they produce quick-and-simple forecasts that can not be too bad. However, the longer the forecast horizon, the worse these models will forecast and they will almost always be out-performed by the more sophisticated, though still parsimonious, *parametric models* that we will start to look at next week.

8.1 Constant level forecast

In the constant level model, it is assumed that the value next period will be the same as this period's value, or formally

$$\widehat{y}_{t+1,t} = y_t.$$

Using the model to forecast ahead for h periods we have simply

$$\widehat{y}_{t+h,t} = y_t.$$

8.2 Constant change forecast

In the constant change model it is assumed that the forecast variable y_t will change by the same amount next period as it changed last period. Formally the model is

$$\Delta\widehat{y}_{t+1,t} = \Delta y_t$$

where Δ is the difference operator defined by $\Delta x_t \equiv x_t - x_{t-1}$. This model can be rewritten, equivalently, as

$$\widehat{y}_{t+1,t} = 2y_t - y_{t-1} = y_t + \Delta y_t.$$

Using the model to forecast ahead for h periods we have

$$\widehat{y}_{t+h,t} = y_t + h\Delta y_t.$$

8.3 Constant growth forecast

In the constant growth model it is assumed that the forecast variable y_t will grow at the same rate next period as it grew last period. Defining the growth rate of y_t as the proportional change in y_t, $\Delta y_t/y_{t-1}$, the model is

$$\frac{\widehat{y}_{t+1,t} - y_t}{y_t} = \frac{y_t - y_{t-1}}{y_{t-1}}$$

or

$$\widehat{y}_{t+1,t} = y_t\frac{y_t}{y_{t-1}}$$

and, forecasting h periods ahead we have

$$\widehat{y}_{t+h,t} = y_t\left(\frac{y_t}{y_{t-1}}\right)^h.$$

Alternatively, defining the growth rate of y_t as the change in the logarithm of y_t, $\Delta\log(y_t)$, the model is

$$\Delta\log(\widehat{y}_{t+1,t}) = \Delta\log(y_t)$$

or

$$\log(\widehat{y}_{t+1,t}) = 2\log(y_t) - \log(y_{t-1}) = \log(y_t) + \Delta\log(y_t)$$

and, forecasting h periods ahead, we have

$$\log(\widehat{y}_{t+h,t}) = \log(y_t) + h\Delta\log(y_t).$$

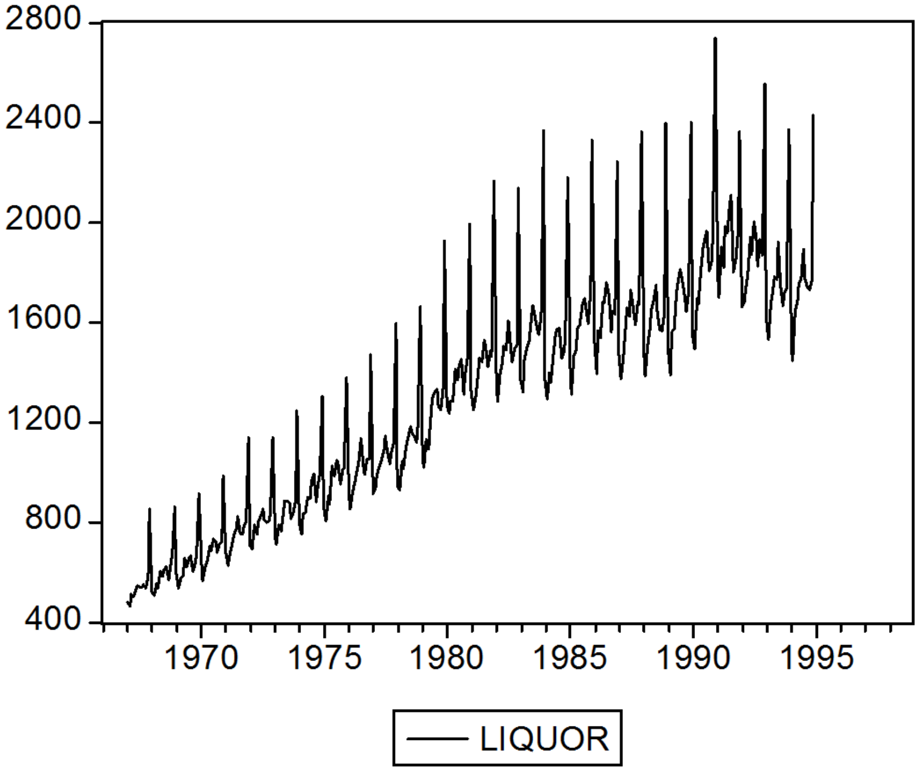

Figure 2: US Liquor Sales: 1967:01–1994:12

9 Decomposing a Time Series Into Components

Looking at graphs of various economic and business time series, we see many common features. Consider three series, taken at random from the datasets analysed in Diebold's textbook and graphed as Figures 2–4. Figure 2 graphs US monthly sales of liquor from 1967–1994. There are two prominent features: a marked upward trend in sales over the whole period and a large spike in sales occuring at the end of each year, corresponding to the Christmas period. This striking seasonal pattern seems to be roughly proportional to the basic trend in the series. A third feature, less obvious than the others is a cyclical pattern, with a period of around four years, roughly corresponding with the US business cycle.

Figure 3 graphs monthly US housing starts from 1946–1994. This series has no obvious trend but a very clear cycle with a period of about seven years. This series also has a strong seasonal pattern with a peak in the spring and the lowest

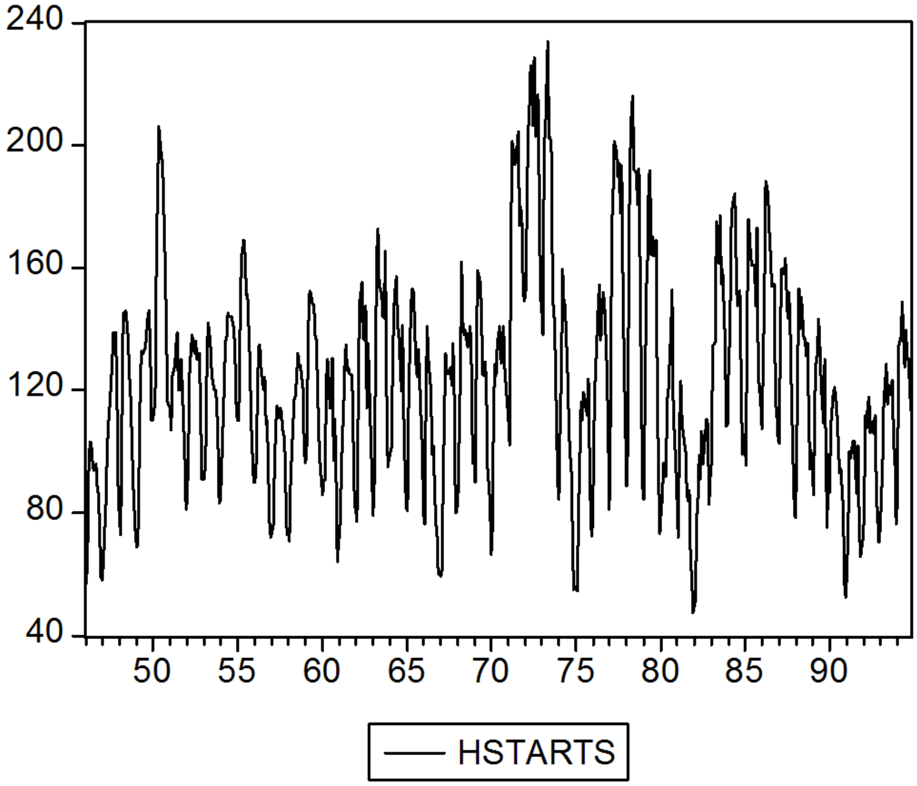

Figure 3: US Housing Starts: 1946:01–1994:11

point in the winter.

Figure 4 graphs the volume of transactions on the New York Stock Exchange, monthly from 1947–1991. This series has a very strong non-linear trend. Around this trend, but dwarfed by it, are cyclical and seasonal patterns, the most striking of which is the big boom around the end of the 1980s.

Together, these three graphs illustrate the main features common to most economic and business time series: a trend, generally (but not always) upwards sloping, cyclical variations corresponding to business and other cycles and (where the variable is observed at monthly or quarterly frequencies) a constant or slowly evolving seasonal pattern within each year.

This empirical observation naturally leads us to consider the theoretical decomposition of a time series into additive components:

$$y_t = T_t + C_t + S_t + u_t$$

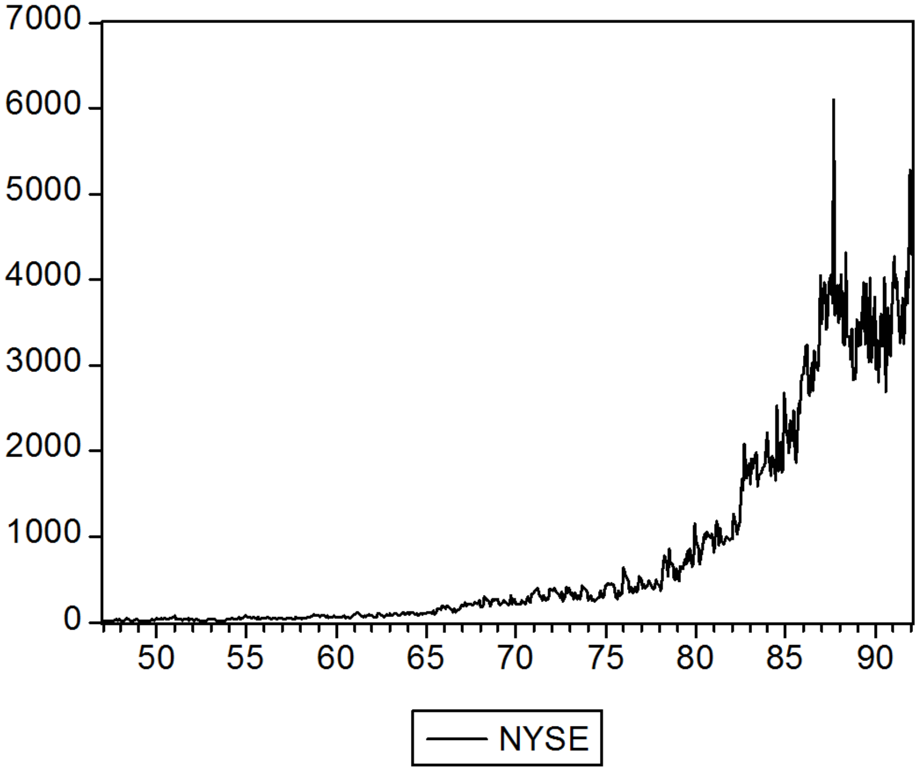

Figure 4: New York Stock Exchange: Volume 1947:01–1992:02

where T_t is the *trend component*, C_t is the *cyclical component* and S_t is the *seasonal component* and u_t is an additional irregular component. An alternative decomposition would be the multiplicative decomposition

$$y_t = T_t^* C_t^* S_t^* u_t^*$$

which gives rise to an additive decomposition in logarithms

$$\log(y_t) = \log(T_t^*) + \log(C_t^*) + \log(S_t^*) + \log(u_t^*).$$

The components in these decompositions are not directly observable. However, for forecasting purposes it is often easier to build models of each component separately and then combine the component forecasts to give a composite forecast:

$$\widehat{y}_{t+1} = \widehat{T}_{t+1} + \widehat{C}_{t+1} + \widehat{S}_{t+1}$$

(where it has been assumed that $\widehat{u}_{t+1} = 0$). This approach will be the method adopted in the next few lectures where we will consider forecasting models for each separate component before putting them all together.

The idea of decomposing a time series into components is adopted more widely than just for forecasting purposes. For example, some economic data series such as unemployment are normally published in a *seasonally adjusted* form. Seasonal adjustment is simply the removal of the seasonal component of a time series and is accomplished by fitting a model of the seasonal pattern to the historical data and subtracting this from the original series as in:

$$y_t^{SA} = y_t - \widehat{S}_t.$$

Similarly, sometimes economic series are *detrended* by removing the trend component as in

$$y_t^{DT} = y_t - \widehat{T}_t.$$

Detrending is often done by economists who wish to remove the trend in order to be able to concentrate on the cyclical properties of a time series. It is widely used in the analysis of *real business cycle* models. Various methods of detrending may be adopted as we will see in next week's lecture. One popular method is the *Hodrick-Prescott* filter which can have the unfortunate property of distorting the cyclical properties of the original series and creating spurious cycles.