Statistical Backgrounds for Forecasting

Terms

+ Forecasts are based on data or observations in the form of a **time series**

+ Suppose there are T **periods** available, with period T being the most recent. The **observation** variable at time period t will be denoted by $y_t, t = 1, 2, \ldots, T$.

+ An observation can be a **cumulative** quantity or an **instantaneous** quantity.

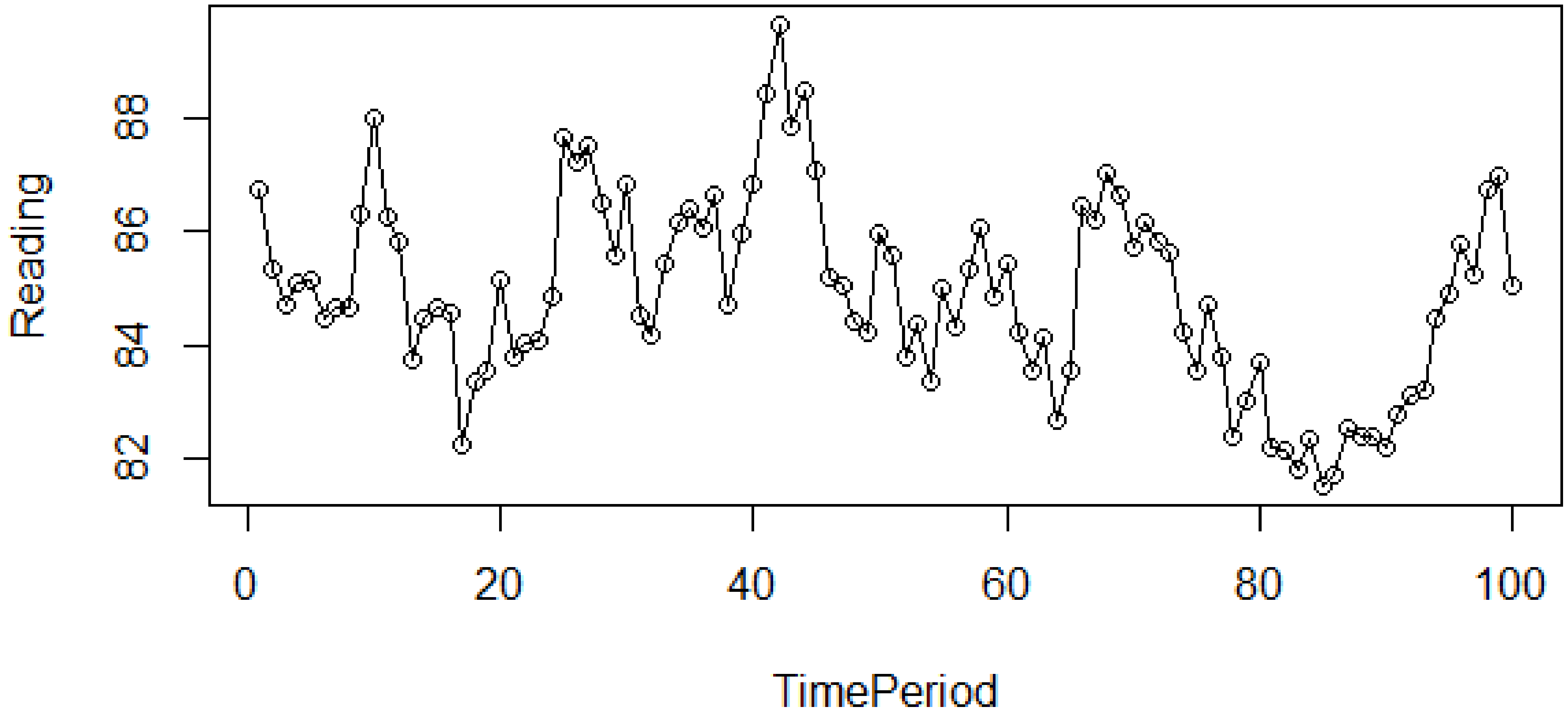
Notations

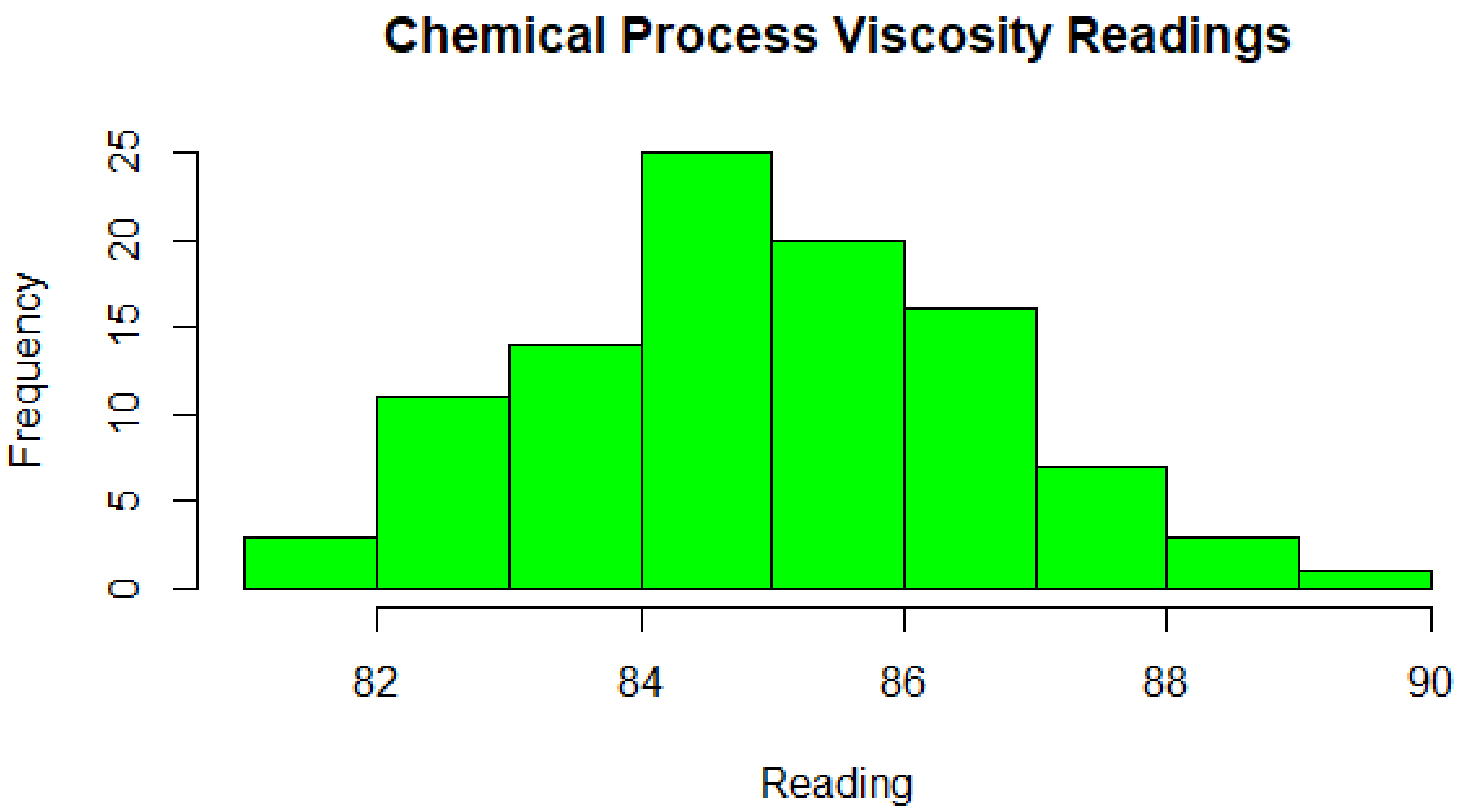
- Let τ be the **forecast lead time**.
- $\hat{y}_t(t-\tau)$ **forecast or predicted value** of y_t
- $\hat{y}_t$ **fitted value** of y_t
- $e_t(\tau) = y_t - \hat{y}_t(t-\tau)$ **lead $\boldsymbol{\tau}$ forecast error**
- $e_t = y_t - \hat{y}_t$ **residual**

- Models usually fit historical data better than they forecast. i.e. $e_t < e_t(\tau)$

Graphical Displays

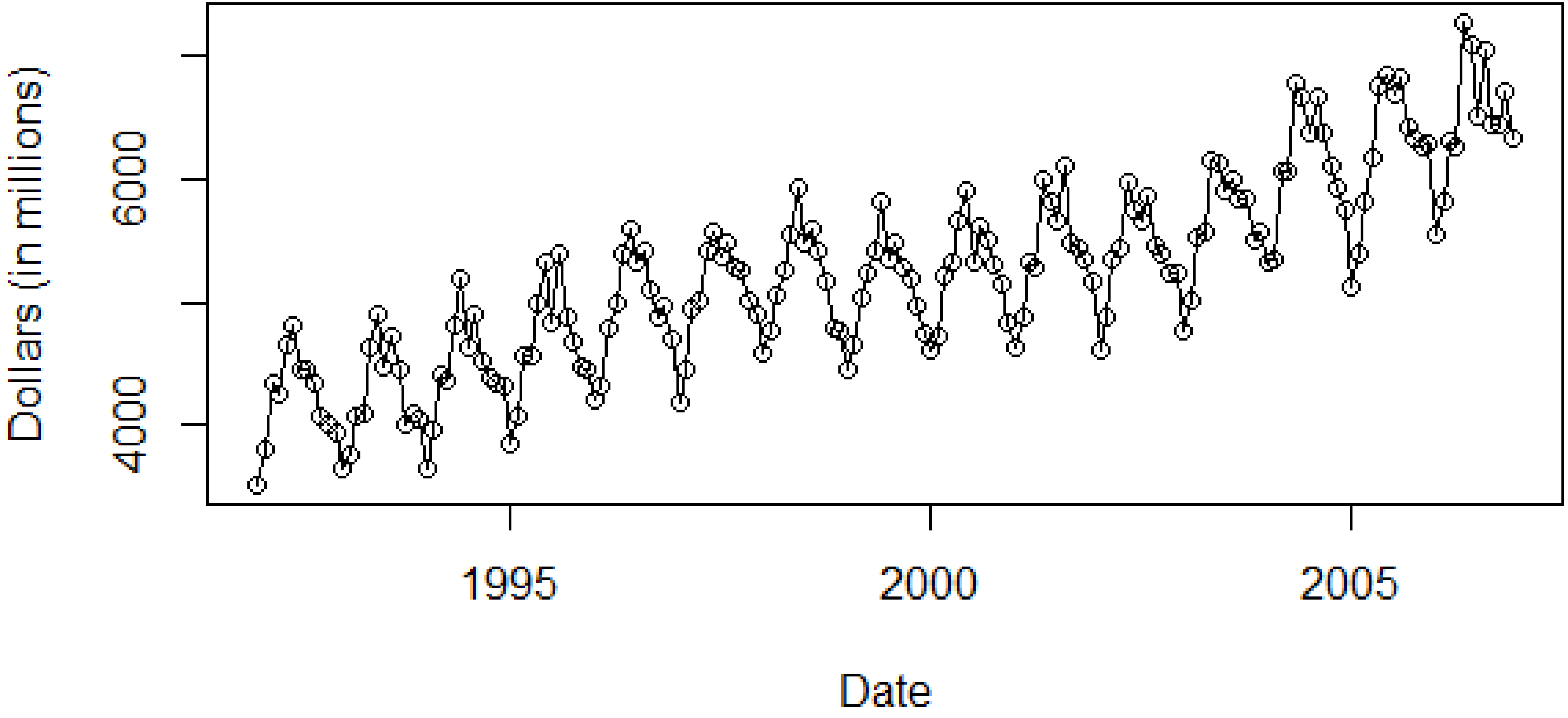
Chemical Process Viscosity Over 100 Periods

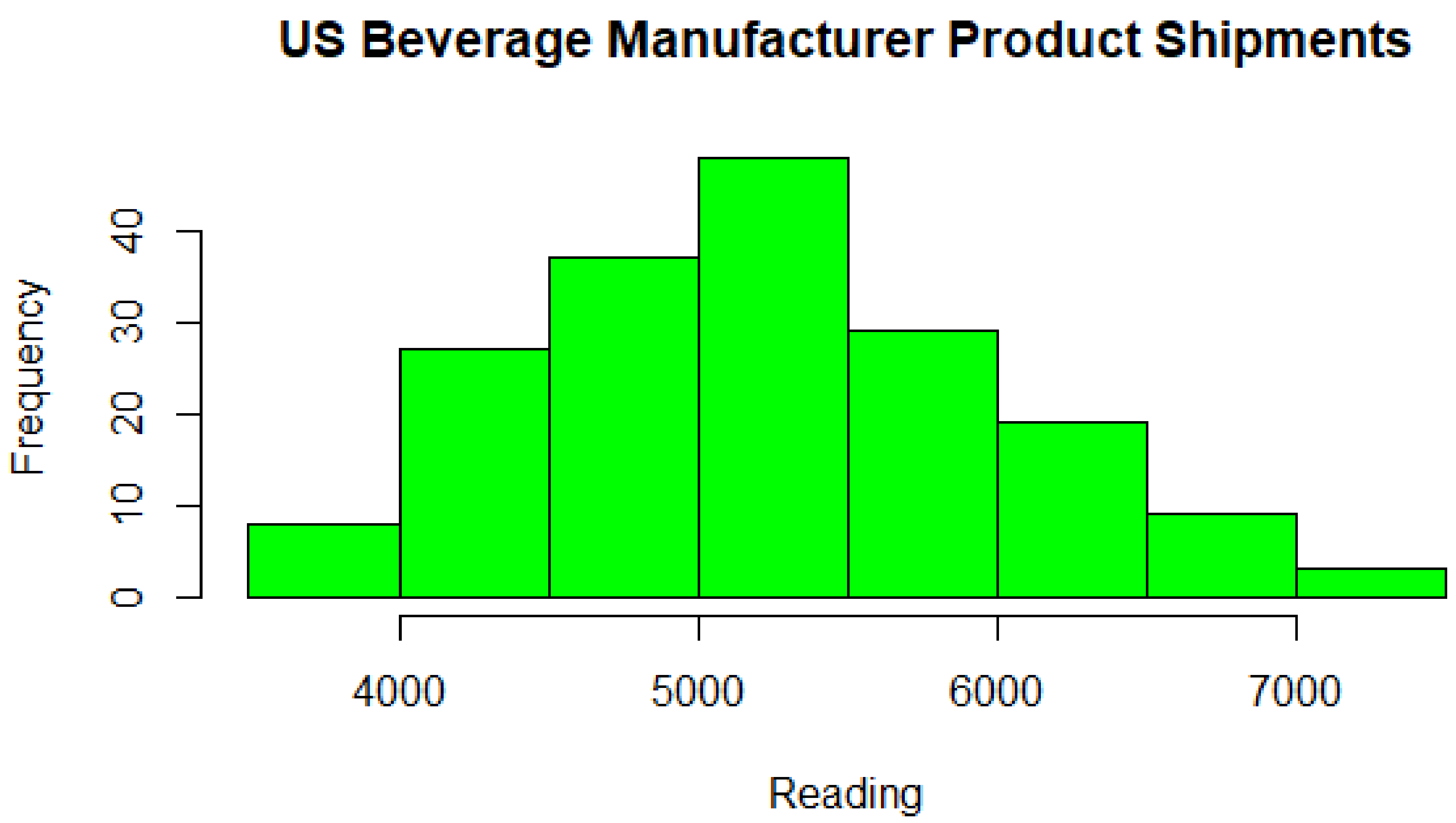


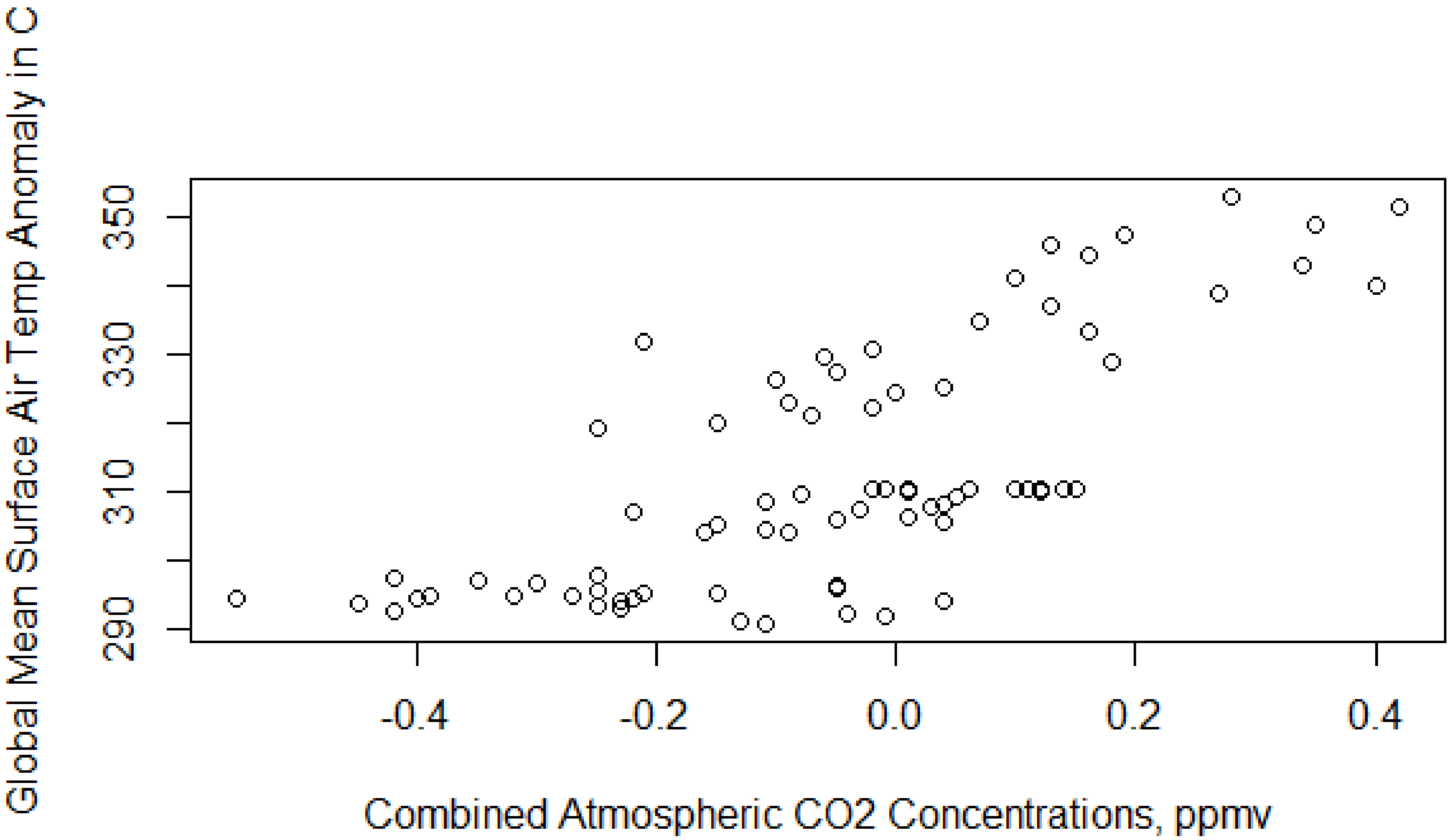


Chemical Process Viscosity Readings
Frequency
0
5
10
15
20
25
82
84
86
88
90
Reading

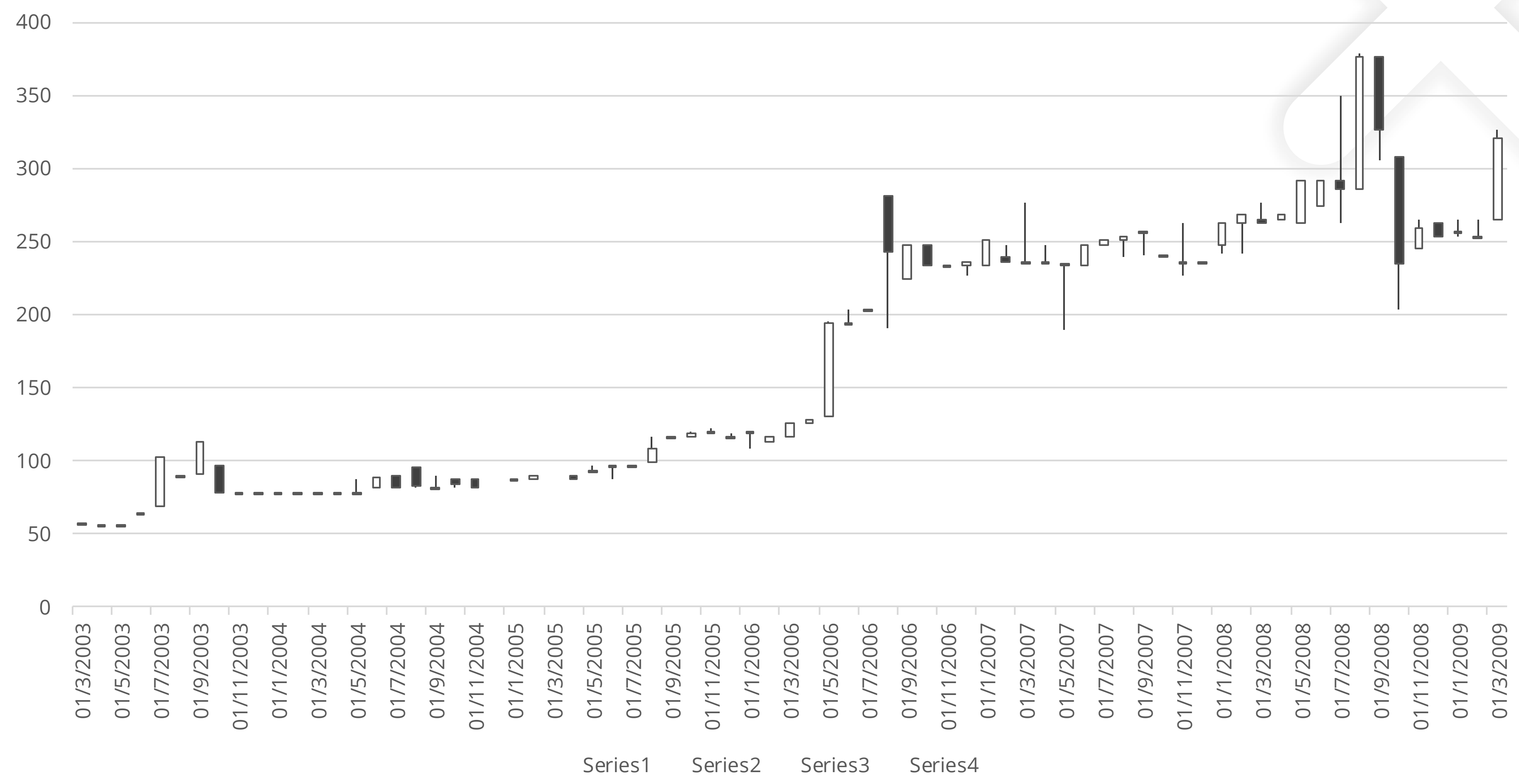
U.S. Beverage Manufacturer Product Monthly Shipments, Unadjusted



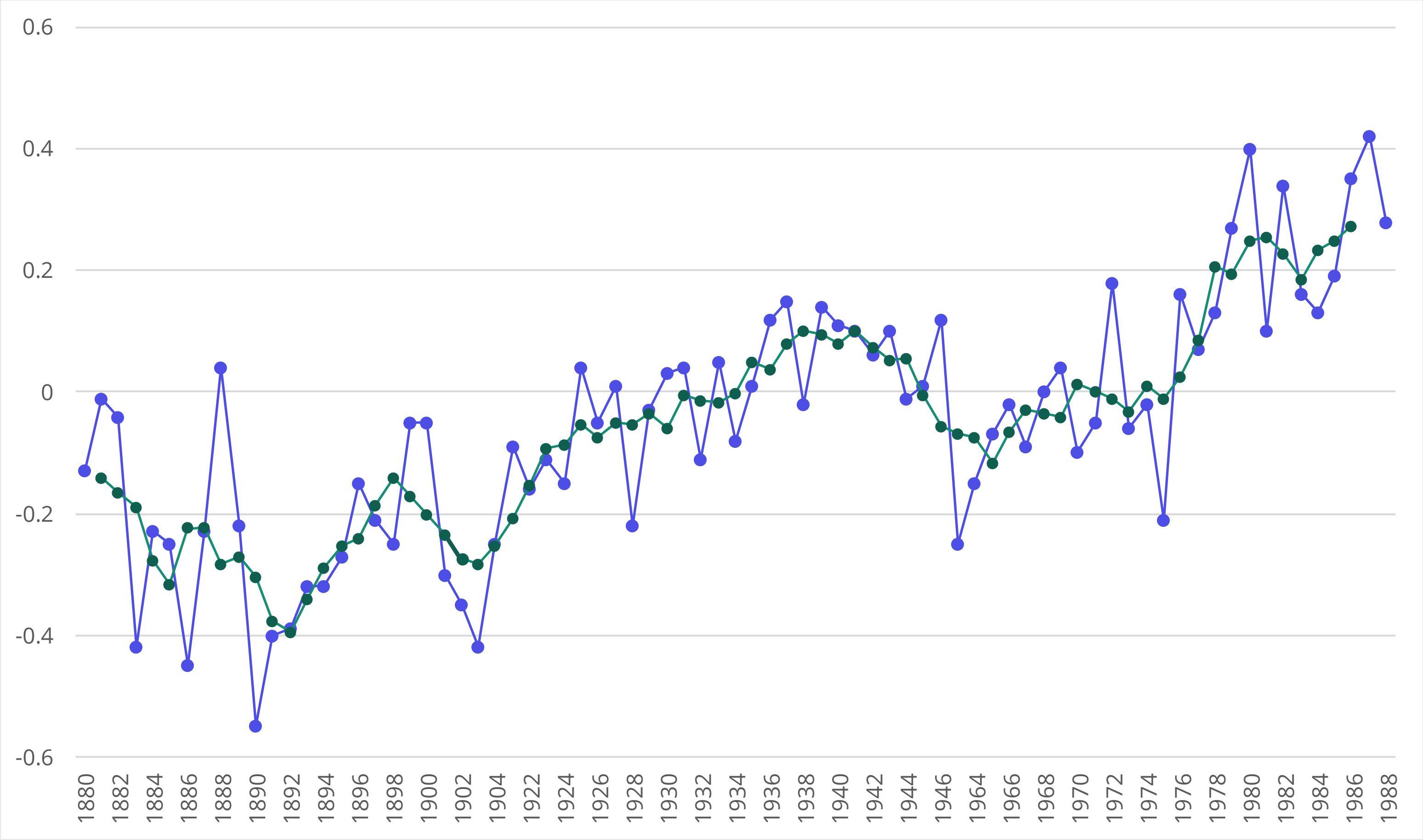

US Beverage Manufacturer Product Shipments
Frequency
0
10
20
30
40
4000
5000
6000
7000
Reading


Global Mean Surface Air Temp Anomaly in C
350
330
310
290
-0.4
-0.2
0.0
0.2
0.4
Combined Atmospheric CO2 Concentrations, ppmv

The scatter plot cannot establish a causal relationship between two variables (neither can naive statistical modeling techniques, such as regression), but it is useful in displaying how the variables have varied together in the historical data set.


400
350
300
250
200
150
100
50
0
01/3/2003
01/5/2003
01/7/2003
01/9/2003
01/11/2003
01/1/2004
01/3/2004
01/5/2004
01/7/2004
01/9/2004
01/11/2004
01/1/2005
01/3/2005
01/5/2005
01/7/2005
01/9/2005
01/11/2005
01/1/2006
01/3/2006
01/5/2006
01/7/2006
01/9/2006
01/11/2006
01/1/2007
01/3/2007
01/5/2007
01/7/2007
01/9/2007
01/11/2007
01/1/2008
01/3/2008
01/5/2008
01/7/2008
01/9/2008
01/11/2008
01/1/2009
01/3/2009
Series1
Series2
Series3
Series4

Smoothed Data



Moving Average

A way to make a smoothed version of the original time series

Moving average of span N

$$M_T = \frac{y_T + y_{T-1} + \cdots + y_{T-N+1}}{N}$$

A moving average have less variability than the original observations.

$$Var(M_T) = Var\left(\frac{1}{N}\sum_{i=T-N+1}^{T} y_i\right) = \frac{1}{N^2}\sum Var(y_t) = \frac{\sigma^2}{N}$$

+ Centered version of moving average of span $2S+1$

$$M_t = \frac{1}{2S+1}\sum_{i=-S}^{S} y_{t+i}$$

+ Hanning Filter

$$M_t^H = 0.25y_{t+1} + 0.5y_t + 0.25y_{t-1}$$

An obvious disadvantage of a linear filter such as a moving average is that an unusual or erroneous data point or an outlier will dominate the averages that contain that observation, contaminating the moving averages for a length of time equal to the span of the filter.

+ Moving medians

$$m_t^{[N]} = med(y_{t-u}, \cdots, y_t, \cdots y_{t+u})$$

where $2u + 1 = N$.

Numerical Description of Time Series Data

Stationary Time Series

- A time series is said to be **strictly stationary** if its properties are not affected by a change in the time origin.

i.e. The joint probability distribution of $(y_t, y_{t+1}, \ldots, y_{t+n})$ is exactly the same as the joint probability distribution of $(y_{t+k}, y_{t+k+1}, \ldots, y_{t+k+n})$

- When n = 0 the stationarity assumption means that the probability distribution of y_t is the same for all time periods and can be written as $f(y)$.

Note that the time series seem to vary around a fixed level. This is a characteristic of stationary time series.
BookB2_Pharmaceutical.mtw

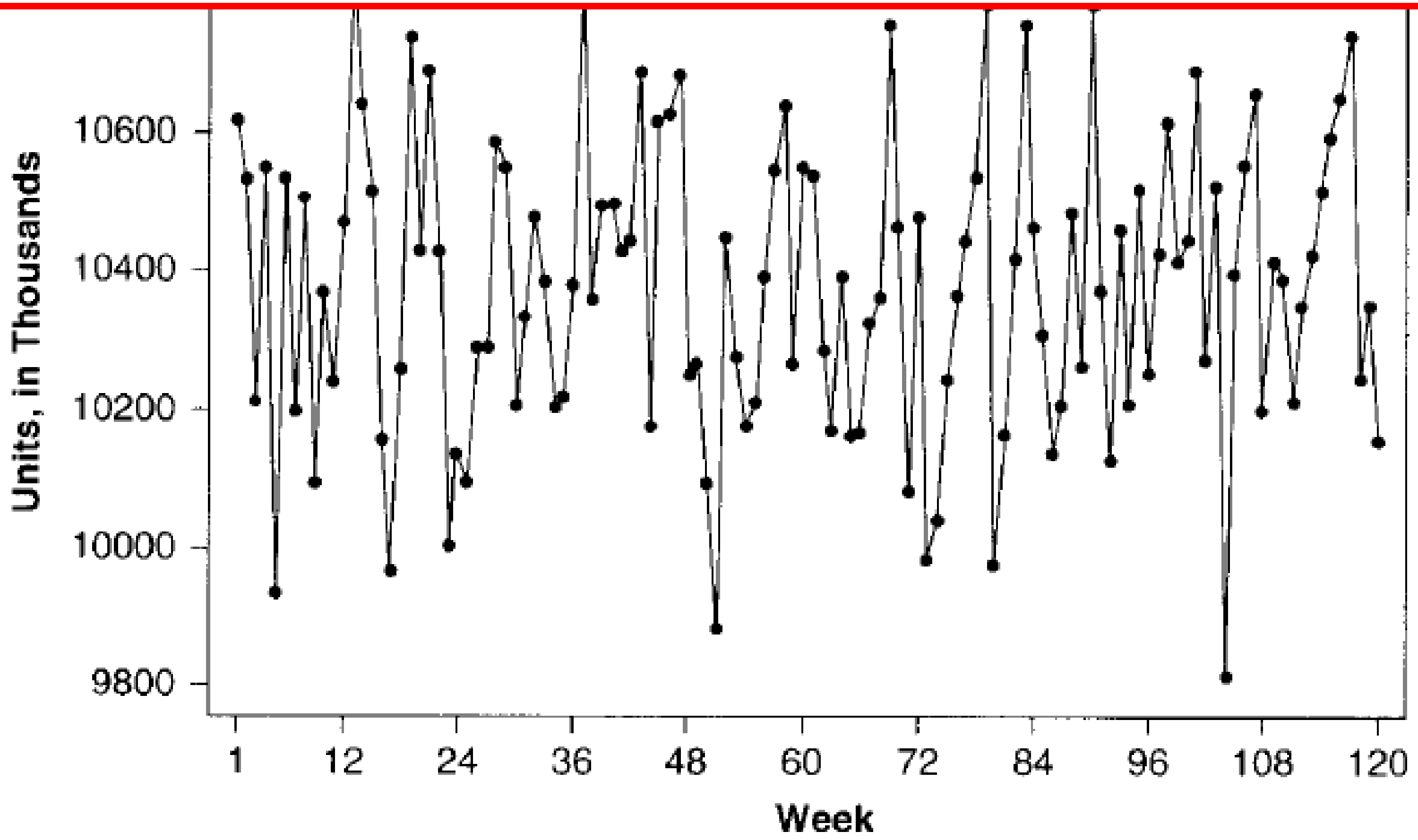


FIGURE 2.8 Pharmaceutical product sales.

Stationary implies a type of statistical equilibrium or stability in the data. Consequently, the time series has

+ Constant mean

$$\mu_y = E(y) = \int_{-\infty}^{\infty} y f(y) dy$$

+ Constant variance

$$\sigma^2 = Var(y) = \int_{-\infty}^{\infty} (y - \mu_y)^2 f(y) dy$$

+ Sample mean

$$\bar{y} = \hat{\mu}_y = \frac{1}{T} \sum_{t=1}^{T} y_t$$

+ Sample variance

$$s^2 = \hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^{T} (y_t - \bar{y})^2$$

- If a time series is stationary this means that the joint probability distribution of any two observations, say, y_t and y_{t+k}. is the same for any two time periods t and t + k that are separated by the same interval k.
- Useful information about this joint distribution. and hence about the nature of the time series, can be obtained by plotting a scatter diagram of all of the data pairs y_t and y_{t+k} that are separated by the same interval k.
- The interval k is called the lag.

The plotted pairs of adjacent observations yt, yt+1 seem to be uncorrelated.
That is, the value of y in the current period does not provide any useful information about the value of y that will be observed in the next period.
BookB2_Pharmaceutical.mtw

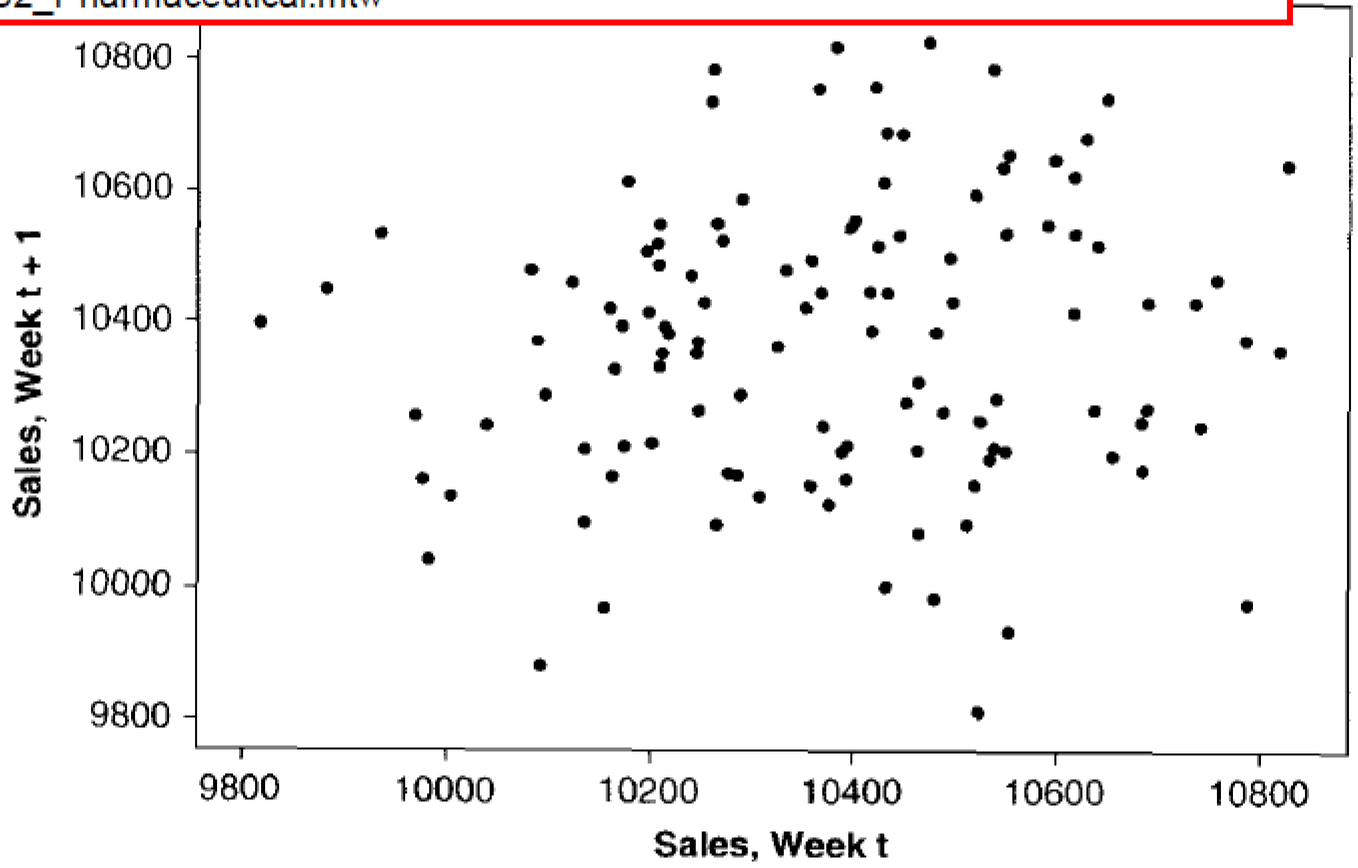


FIGURE 2.10 Scatter diagram of pharmaceutical product sales at lag $k = 1$.

The pairs of adjacent observations yt+l, y, are positively correlated. That is, a small value of y tends to be followed in the next time period by another small value of y. and a large value of y tends to be followed immediately by another large value of y.
BookB3_Viscosity.mtw

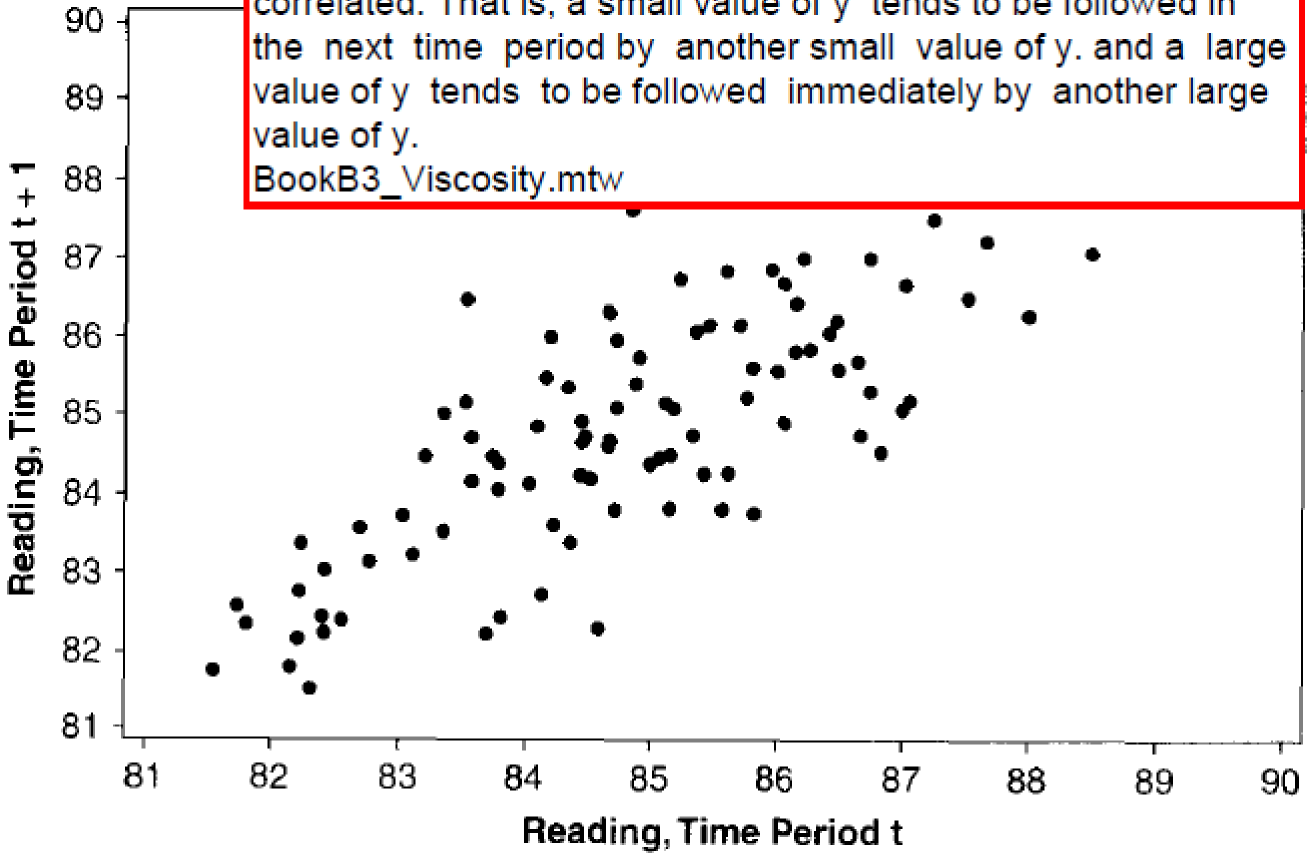


FIGURE 2.11 Scatter diagram of chemical viscosity readings at lag $k = 1$.

Autocovariance at lag k

$$\gamma_k = Cov(y_t, y_{t+k}) = E[(y_t - \mu)(y_{t+k} - \mu)]$$

The collection of the values of $\gamma_k, k = 0, I, 2, \ldots$ is called the **autocovariance function.**

Note that the autocovariance at lag k = 0 is just the variance of the time series.

Autocorrelation Coefficient at lag k

$$\rho_k = \frac{\gamma_k}{\gamma_0}$$

Autocorrelation Coefficient at lag k

$$\rho_k = \frac{\gamma_k}{\gamma_0}$$

The collection of the values of $\rho_k. k = 0. I. 2 \ ....$ is called the **autocorrelation function (ACF).**

ACF is independent of the scale of measurement of the time series, so it is a dimensionless quantity. Furthermore. $\rho_k = \rho_{-k}$ that is, the autocorrelation function is **symmetric around zero.** so it is only necessary to compute the positive (or negative) half.

If a time series has a finite mean and autocovariance function it is said to be **second order stationary (or weakly stationary of order 2).**

If. in addition. the joint probability distribution of the observations at all times is multivariate normal. then that would be sufficient to result in a time series that is strictly stationary.

+ Sample autocovariance

$$c_k = \hat{\gamma}_k = \frac{1}{T}\sum_{t=1}^{T-k}(y_t - \mu)(y_{t+k} - \mu)$$

+ Sample ACF

$$r_k = \hat{\rho}_k = \frac{c_k}{c_0}$$

A good general rule of thumb is that at least 50 observations are required to give a reliable estimate of the ACF, and the individual sample autocorrelations should be calculated up to lag K. where K is about T/4.

Note the rate of decrease or decay in ACF values from 0. 78 to 0, followed by a sinusoidal pattern about 0. This ACF pattern is typical of stationary time series. The plot of sample ACFs for a time series of random values with constant mean has a much different appearance.

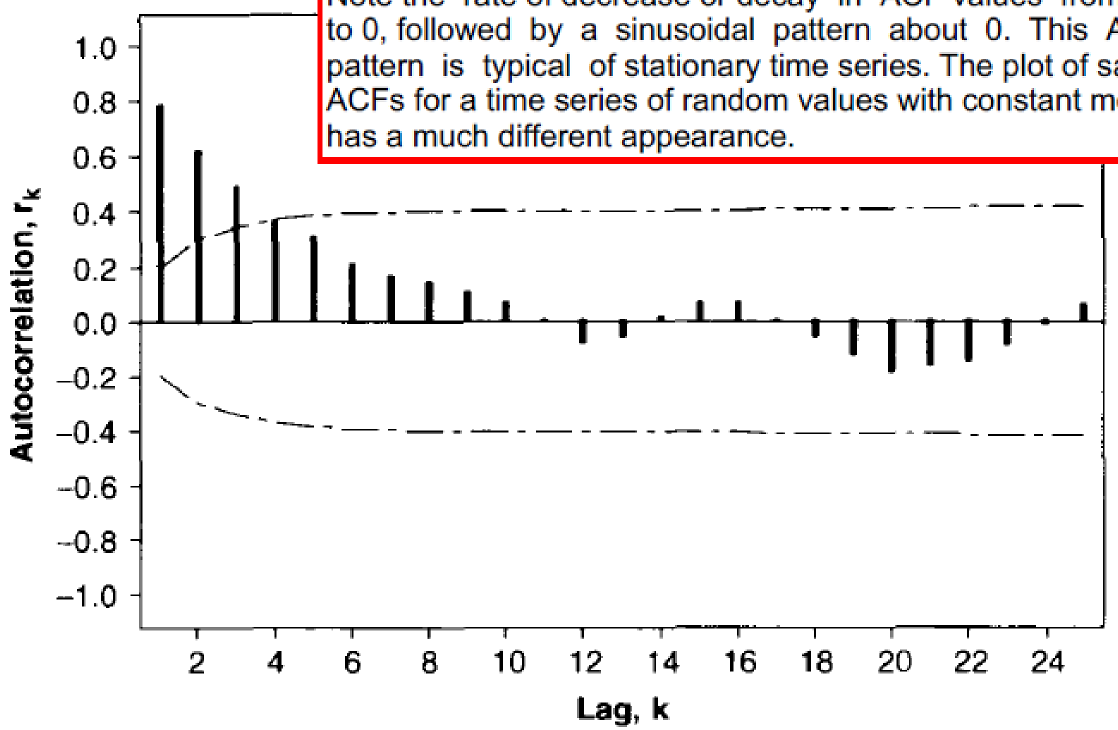


FIGURE 2.12 Sample autocorrelation function for chemical viscosity readings, with 5% significance limits.

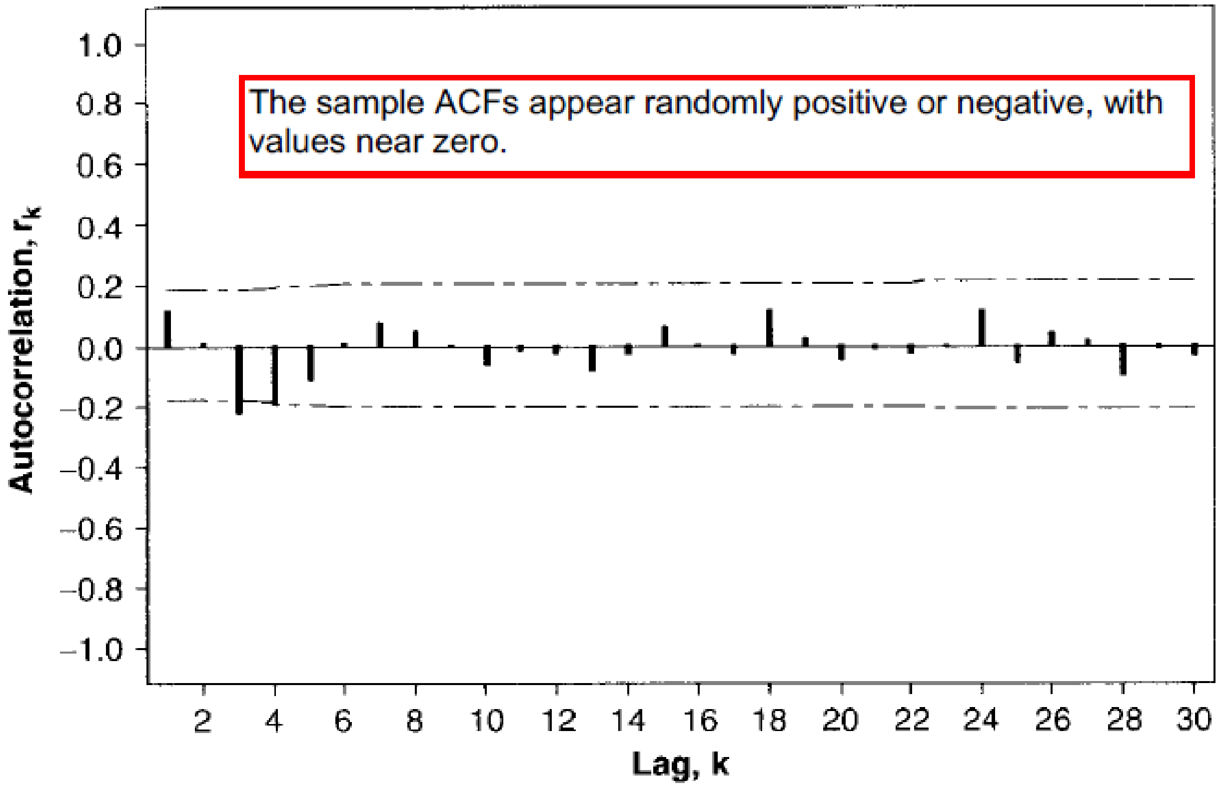
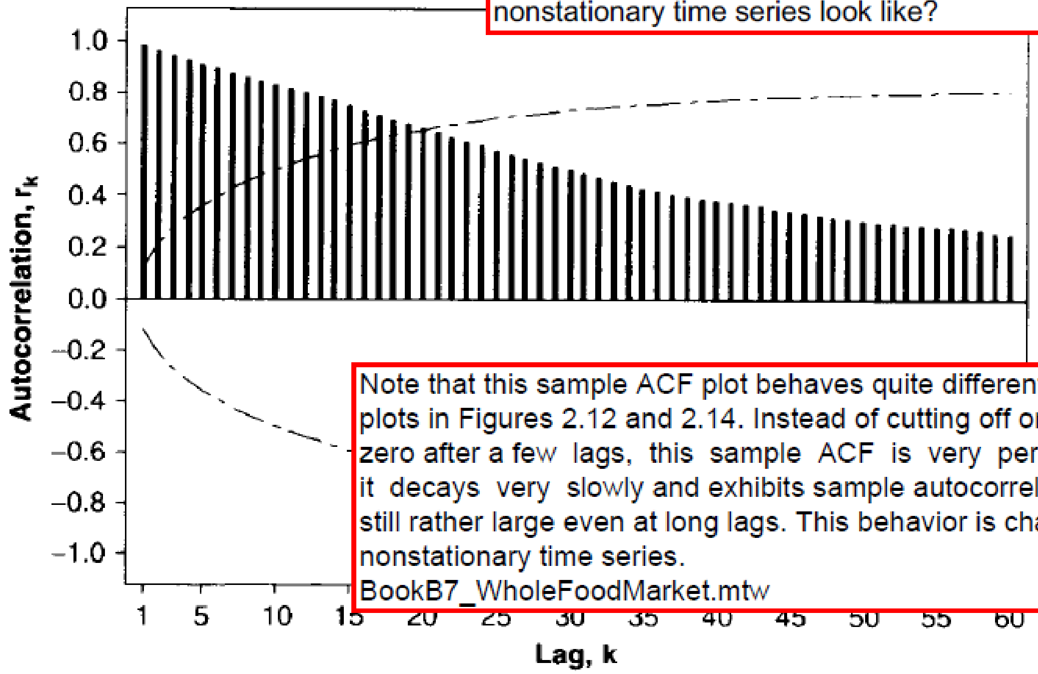


FIGURE 2.14 Autocorrelation function for pharmaceutical product sales, with 5% significance limits.

While the autocorrelation function is defined only for a stationary time series (restrições no cálculo da variância) the sample ACF can be computed for any time series, so a logical question is: What does the sample ACF of a nonstationary time series look like?



Note that this sample ACF plot behaves quite differently than the ACF plots in Figures 2.12 and 2.14. Instead of cutting off or tailing off near zero after a few lags, this sample ACF is very persistent; that is, it decays very slowly and exhibits sample autocorrelations that are still rather large even at long lags. This behavior is characteristic of a nonstationary time series.

BookB7_WholeFoodMarket.mtw

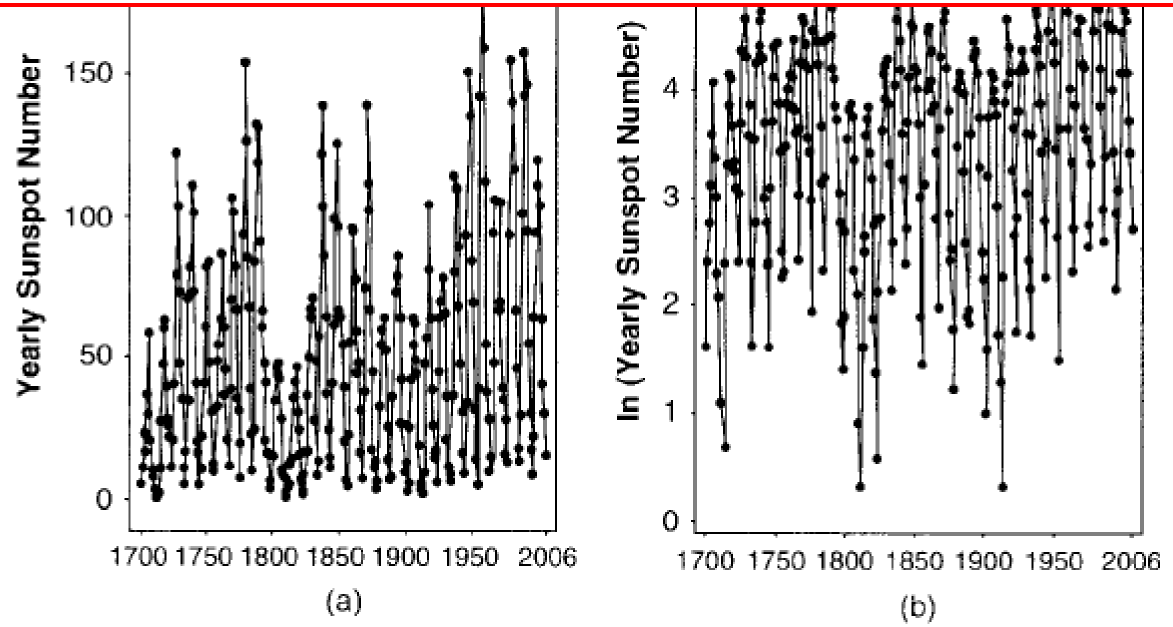
FIGURE 2.15 Autocorrelation function for Whole Foods market stock price, with 5% significance limits.

Data Transformations and Adjustments

- Data transformation are often used for stabilizing the variance of data.
- Non-constant variance is common for time series data

Figure shows cyclic patterns of varying magnitudes. The variability from about 1800 to 1830 is smaller than that from about 1830 to 1880; other small periods of constant, but different variances can also be identified.
BookB9_SunspotNumber.mtw

Faça usando o Minitab
Veja BoxCox.ppt



The application of a natural logarithm transformation tends to stabilize the variance and leaves just a few unusual values.

Power family of transformations

$$y^{(\lambda)} = \begin{cases} \dfrac{y^{\lambda} - 1}{\lambda \dot{y}^{\lambda-1}}, & \lambda \neq 0 \\ \dot{y} \ln y\,, & \lambda = 0 \end{cases}$$

where $\dot{y} = \exp[\frac{1}{T}\sum_{t=1}^{T} y_t]$ is the geometric mean of the observations.

When $\lambda = 1$ there are no transformations.

$\lambda = 0.5$	square root transformation
$\lambda = 0$	log transformation
$\lambda = -0.5$	reciprocal square root transformation
$\lambda = -1$	inverse transformation

- The log transformation is used frequently in situations where the variability in the original time series increases with the average level of the series.
- When the standard deviation of the original series increases linearly with the mean, the log transformation is in fact an optimal variance-stabilizing transformation.

- A time series that exhibits a trend is a nonstationary time series. Modeling and forecasting of such a time series is greatly simplified if we can eliminate the trend.
- One way to do this is to fit a regression model describing the trend component to the data and then subtracting it out of the original observations, leaving a set of residuals that are free of trend

- Linear trend

$$E(y_t) = \beta_0 + \beta_1 t$$

- Quadratic

$$E(y_t) = \beta_0 + \beta_1 t + \beta_2 t^2$$

- Exponential

$$E(y_t) = \beta_0 e^{\beta_1 t}$$

There is clearly a positive nearly linear trend. The trend analysis plot shows the original time series with the fitted

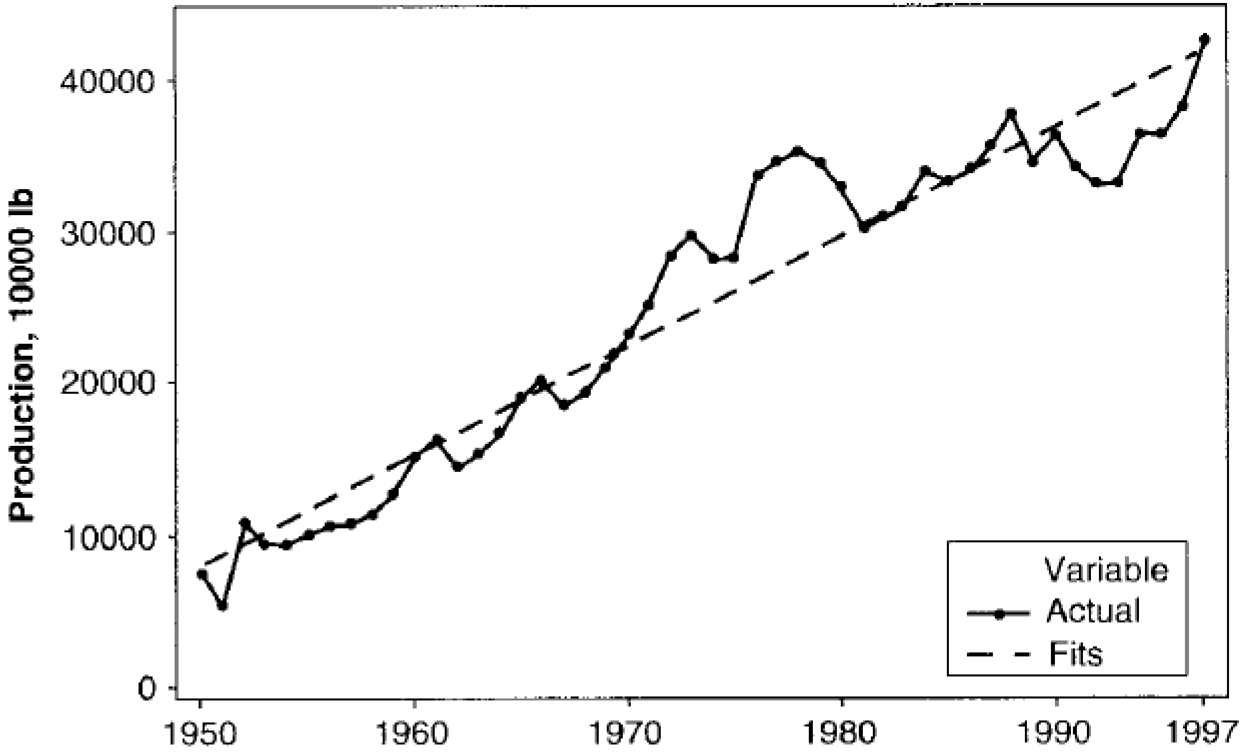
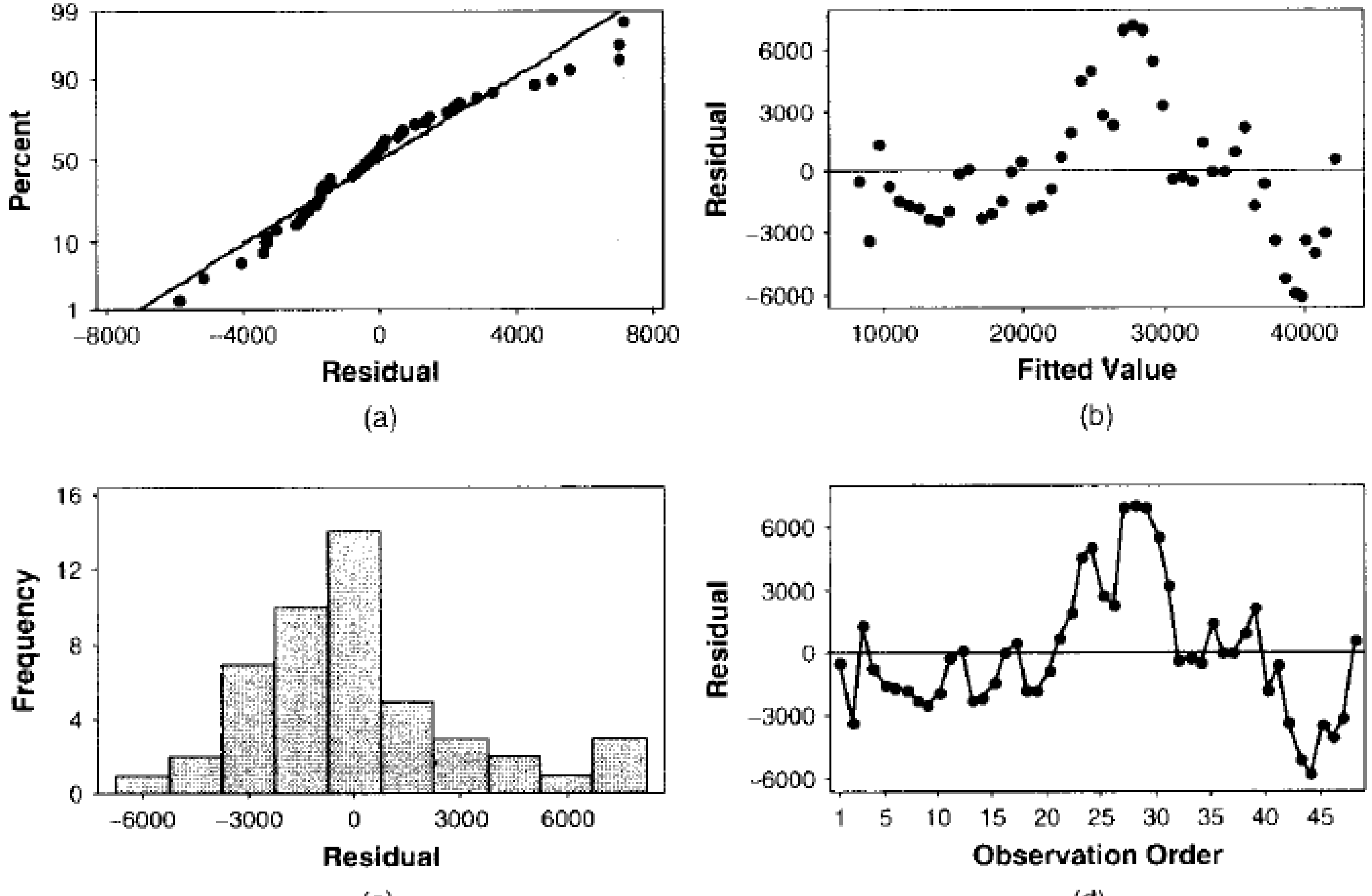


FIGURE 2.17 Blue and gorgonzola cheese production, with fitted regression line. (*Source*: USDA–NASS.)



Plots of the residuals from this model indicate that, in addition to an underlying trend, there is additional structure. The normal probability plot and histogram indicate the residuals are approximately normally distributed.

However, the plots of residuals versus fitted values and versus observation order indicate nonconstant variance in the last half of the time series.

+ Another approach to removing trend is by differencing the data; that is, applying the difference operator to the original time series to obtain a new time series.

$$x_t = y_t - y_{t-1} = \Delta y$$

where Δ is the backward difference operator.

$$x_t = (1 - B)y_t$$

where B is the backshift operator defined by $By_t = y_{t-1}$

$$\mathrm{B}^d y_t = y_{t-d}$$
$$\Delta^d y_t = (1 - B)^d y_t$$

Differencing has two advantages relative to fitting a trend model to the data.

- it does not require estimation of any parameters. so it is a more **parsimonious** (i.e ..simpler) approach; and
- second, model fitting assumes that the trend is fixed through out the time series history and will remain so in the (at least immediate) future. In other words, the trend component, once estimated. is assumed to be **deterministic.** Differencing can allow the trend component to change through time.

- The first difference accounts for a trend that impacts the change in the mean of the time series.
- the second difference accounts for changes in the slope of the time series.
- and so forth. Usually, one or two differences are all that is required in practice to remove an underlying trend in the data.

A difference of one applied to this time series removes the increasing trend and also improves the appearance of the residuals plotted versus fitted value and observation order. This illustrates that differencing may be a very good alternative to detrending a time series by using a regression modeL

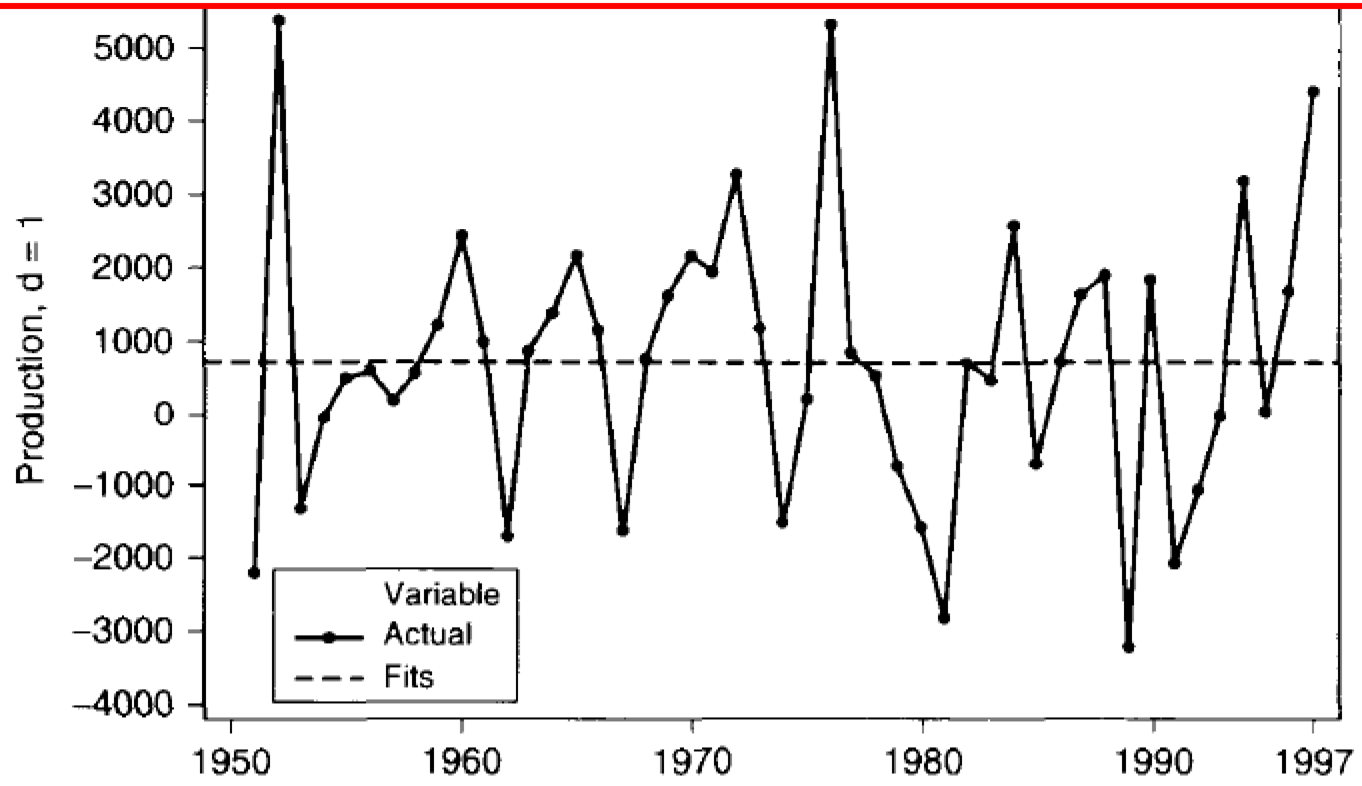


FIGURE 2.19 Blue and gorgonzola cheese production, with one difference. (*Source*: USDA-NASS.)

Differencing is used to simplify the correlation structure and to help reveal any underlying pattern.

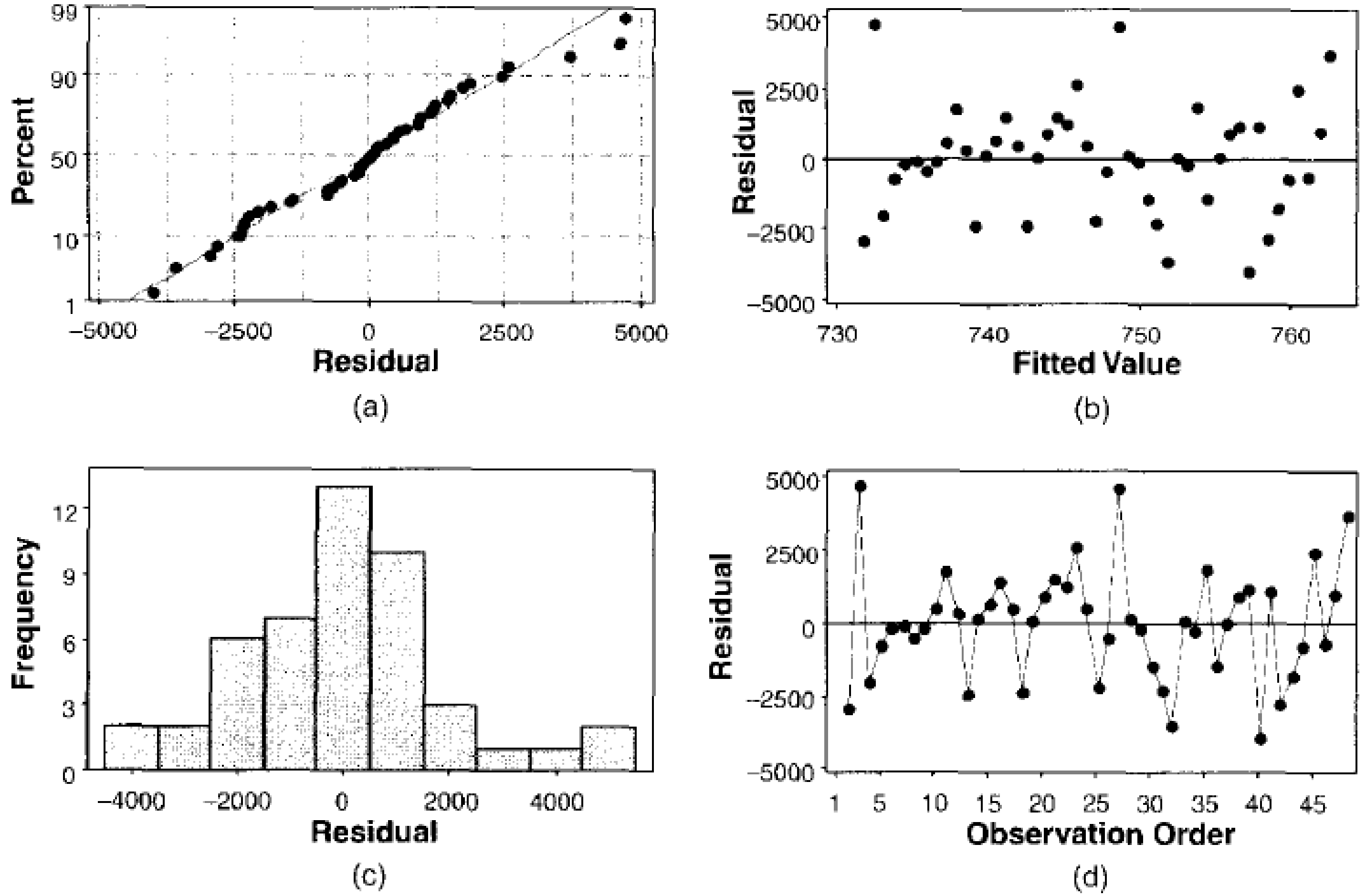


FIGURE 2.20 Residual plots for one difference of blue and gorgonzola cheese production.

- Differencing can also be used to eliminate seasonality. Define a lag-d seasonal difference operator as

$$\Delta_d y_t = (1 - B^d) y_t = y_t - y_{t-d}$$

- When both trend and seasonal components are simultaneously present, we can sequentially difference to eliminate these effects. That is, first seasonally difference to remove the seasonal component and then difference one or more times using the regular difference operator to remove the trend.

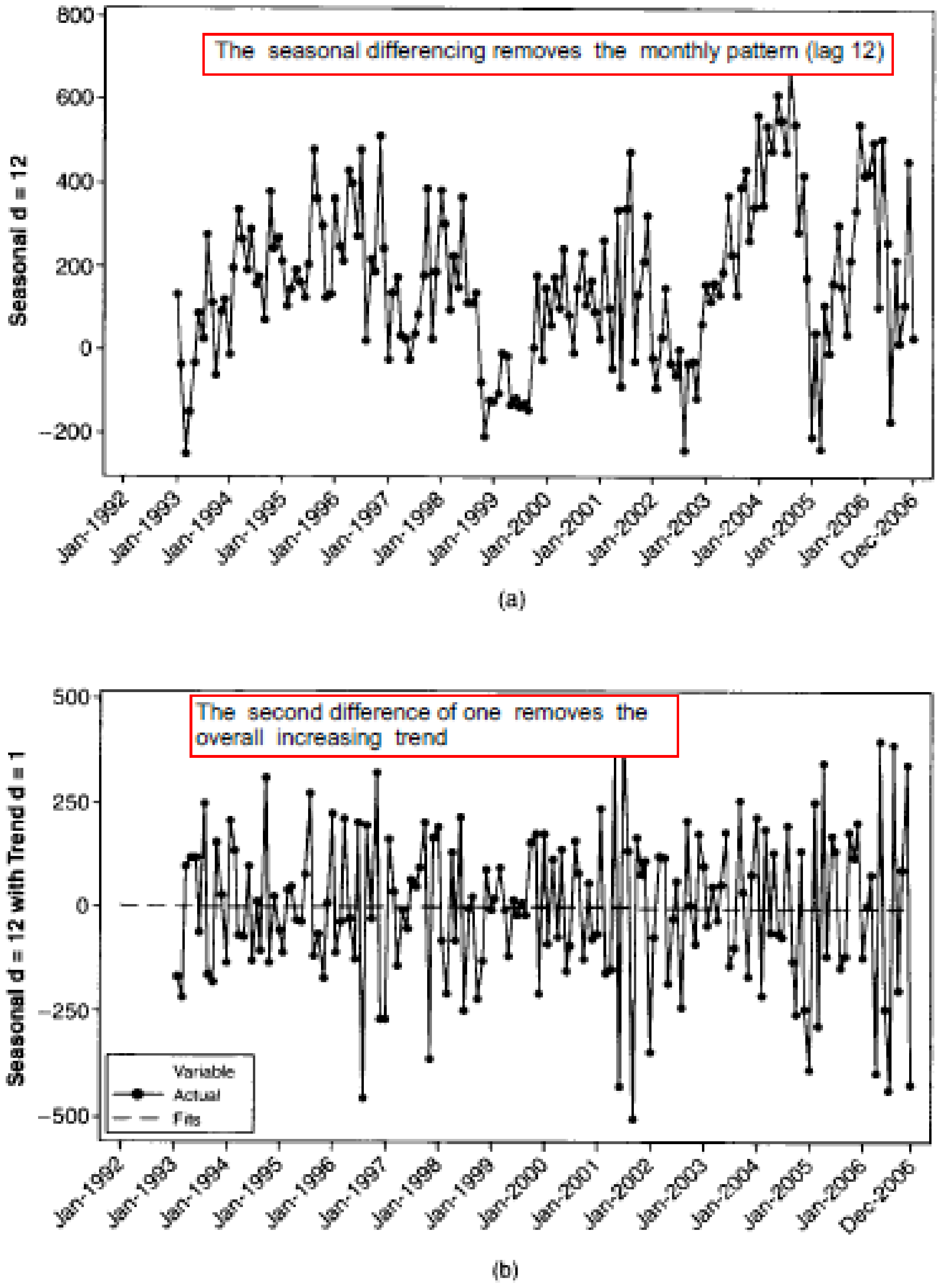


FIGURE 2.21 Time series plots of seasonal- and trend-differenced beverage data.

Examination of the residual plots does not reveal any problems with the linear trend model fit to the differenced data.

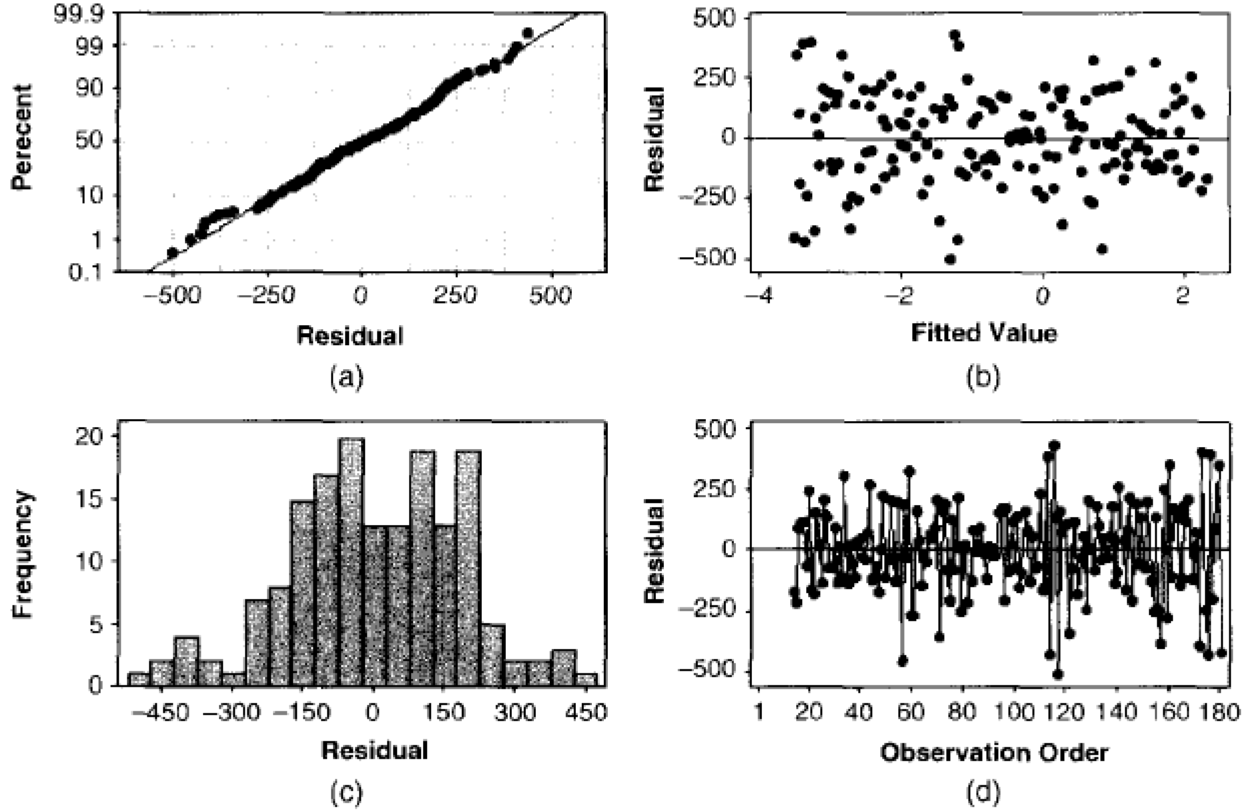


FIGURE 2.22 Residual plots for linear trend model of differenced beverage shipments.

The original data (labeled "Actual") along with the fitted trend line ('Trend") and the predicted values ("Fits") from the additive model with both the trend and seasonal components.

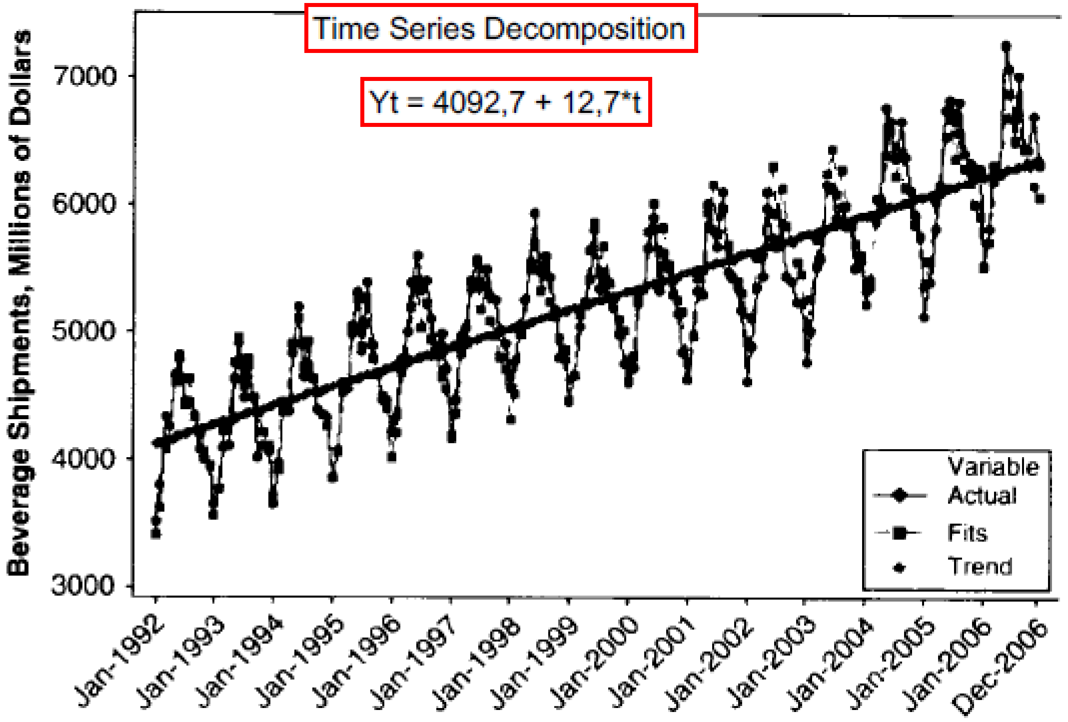


FIGURE 2.23 Time series plot of decomposition model for beverage shipments.

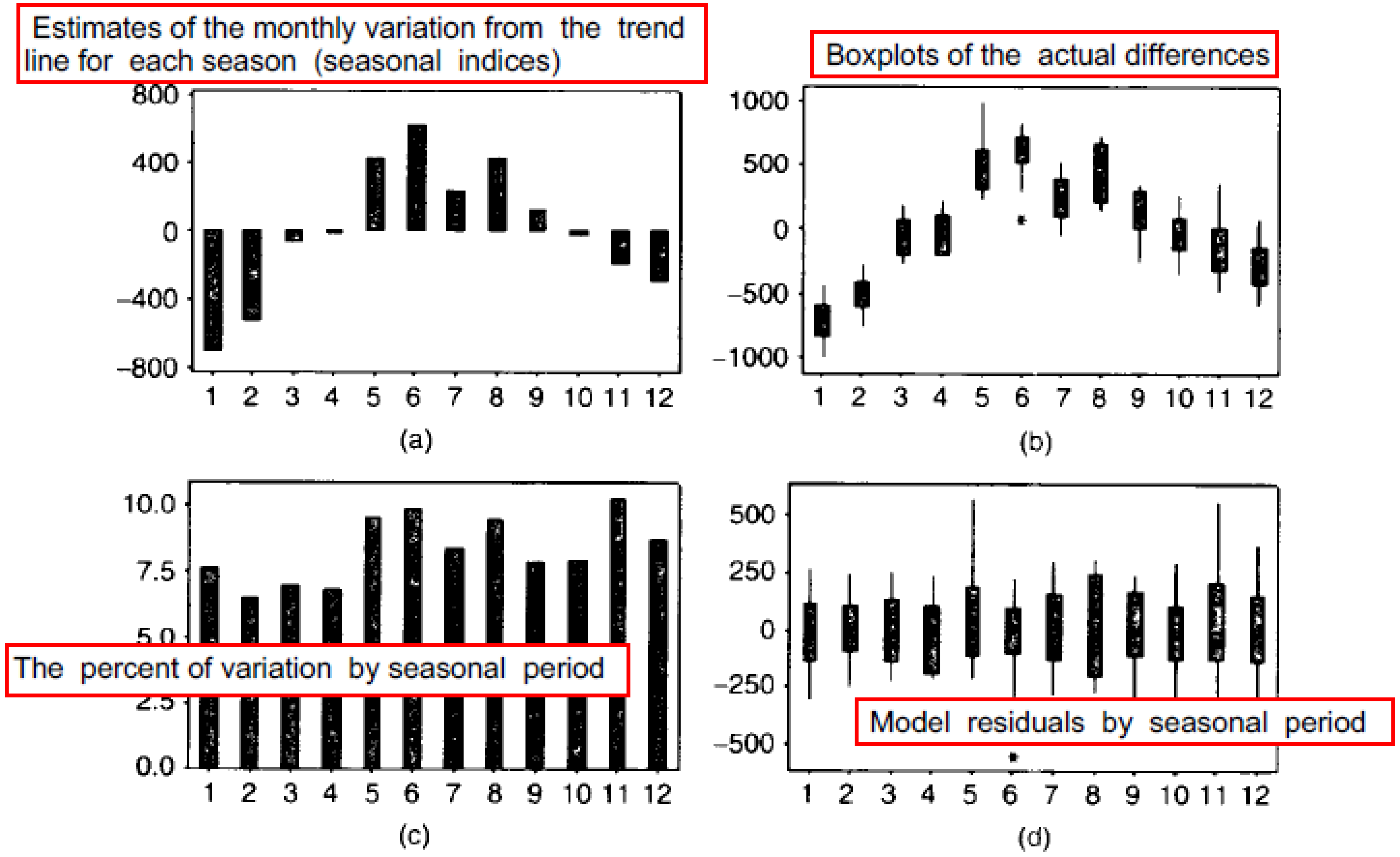


FIGURE 2.24 Seasonal analysis for beverage shipments.

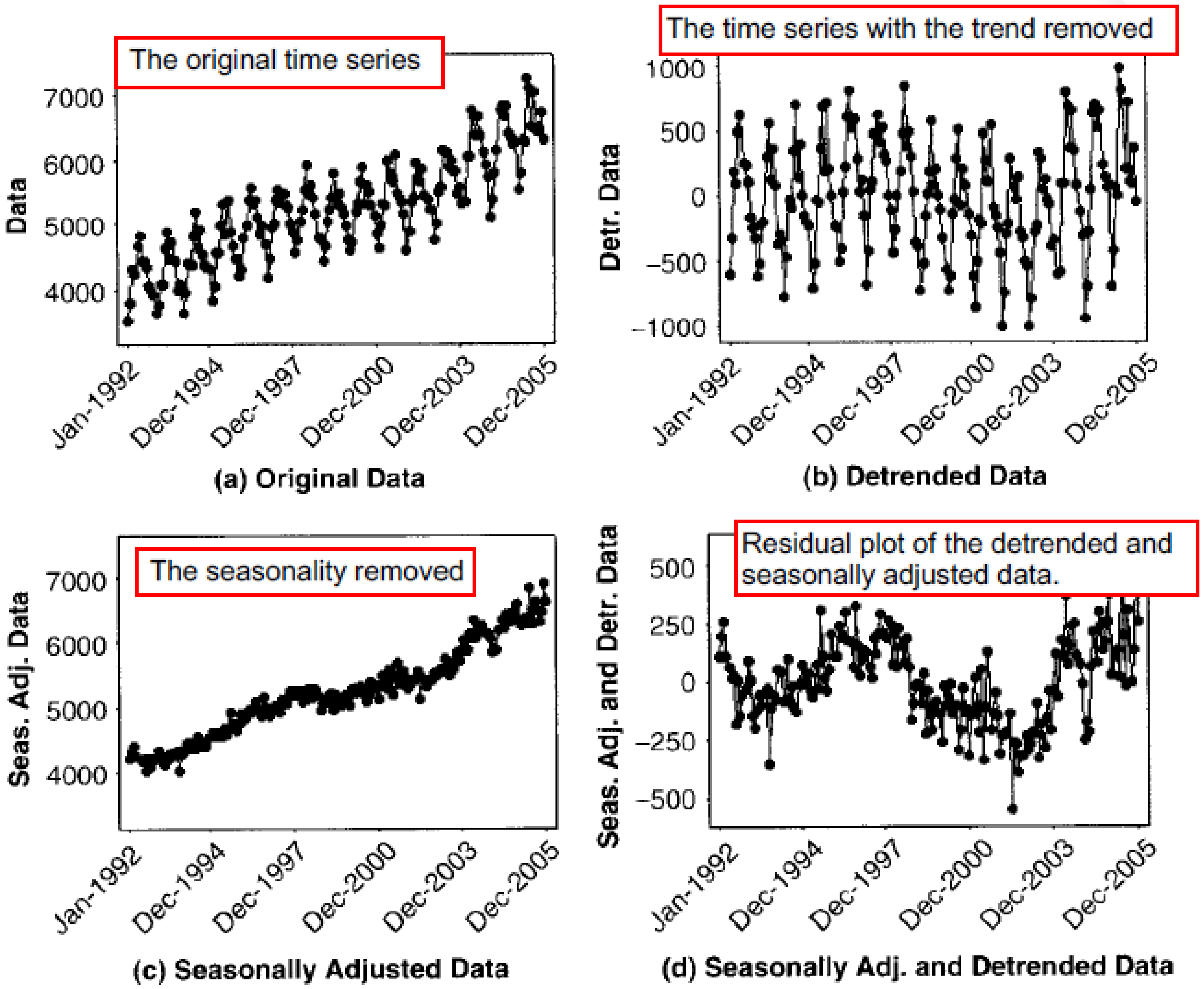


FIGURE 2.25 Component analysis of beverage shipments.

GENERAL APPROACH TO TIME SERIES MODELING AND FORECASTING

Basic Steps

1. Plot the time series and determine its basic features
2. Eliminate any trend or seasonal components
3. Develop a forecasting model for the residuals
4. Validate the performance of the model
5. Also of interest are the differences between the original time series and the values that would be forecast by the model in the original scale
6. For forecasts of future values on the original scale, reverse the transformations used.
7. If prediction intervals are desired for the forecast, construct prediction intervals for the residuals and then reverse the transformations made to produce the residuals as described earlier.
8. Monitor the forecast

Evaluating and monitoring forecasting model performance

Forecasting Model Evaluation

Scale-dependent measures of forecast accuracy

- One-step-ahead forecast error

$$e_t(1) = y_t - \hat{y}_t(t-1)$$

- Average error or mean error

$$ME = \frac{1}{n}\sum_{t=1}^{n} e_t(1)$$

- Mean absolute deviation

$$MAD = \frac{1}{n}\sum_{t=1}^{n} |e_t(1)|$$

- Mean squared error

$$MSE = \frac{1}{n}\sum_{t=1}^{n} [e_t(1)]\hat{}2$$

+ MAD and MSE measure variability of forecast errors

$$\hat{\sigma}^2_{e(1)} = MSE$$

$$\hat{\sigma}_{e(1)} = \sqrt{\frac{\pi}{2}}MAD$$

Relative measures

+ Relative/percent forecast error

$$re_t(1) = \frac{e_t(1)}{y_t} \times 100\%$$

+ Mean percent forecast error

$$ME = \frac{1}{n}\sum_{t=1}^{n} re_t(1)$$

+ Mean absolute percent forecast error

$$MAD = \frac{1}{n}\sum_{t=1}^{n} |re_t(1)|$$

TABLE 2.2 Calculation of Forecast Accuracy Measures

Time Period	(1) Observed Value y_t	(2) Forecast $\hat{y}_t(t-1)$	(3) Forecast Error $e_t(1)$	(4) Absolute Error $\lvert e_t(1)\rvert$	(5) Squared Error $[e_t(1)]^2$	(6) Relative (%) Error $(e_t(1)/y_t)100$	(6) Absolute (%) Error $\lvert(e_t(1)/y_t)100\rvert$
1	47	51.1	−4.1	4.1	16.81	−8.7234	8.723404
2	46	52.9	−6.9	6.9	47.61	−15	15
3	51	48.8	2.2	2.2	4.84	4.313725	4.313725
4	44	48.1	−4.1	4.1	16.81	−9.31818	9.318182
5	54	49.7	4.3	4.3	18.49	7.962963	7.962963
6	47	47.5	−0.5	0.5	0.25	−1.06383	1.06383
7	52	51.2	0.8	0.8	0.64	1.538462	1.538462
8	45	53.1	−8.1	8.1	65.61	−18	18
9	50	54.4	−4.4	4.4	19.36	−8.8	8.8
0	51	51.2	−0.2	0.2	0.04	−0.39216	0.392157
11	49	53.3	−4.3	4.3	18.49	−8.77551	8.77551
12	41	46.5	−5.5	5.5	30.25	−13.4146	13.41463
13	48	53.1	−5.1	5.1	26.01	−10.625	10.625
14	50	52.1	−2.1	2.1	4.41	−4.2	4.2
15	51	46.8	4.2	4.2	17.64	8.235294	8.235294
16	55	47.7	7.3	7.3	53.29	13.27273	13.27273
17	52	45.4	6.6	6.6	43.56	12.69231	12.69231
18	53	47.1	5.9	5.9	34.81	11.13208	11.13208
19	48	51.8	−3.8	3.8	14.44	−7.91667	7.916667
20	52	45.8	6.2	6.2	38.44	11.92308	11.92308
		Totals	*−11.6*	*86.6*	*471.8*	*−35.1588*	*177.3*

The P-value is 0.088, so the hypothesis of normality of the forecast errors would not be rejected at the 0.05 level. Minitab also reports the standard deviation of the forecast errors to be 4.947, a slightly larger value than we computed from the MSE, because Minitab uses the standard method for calculating sample standard deviations.

Reproduza no Minitab.

Arquivo Error.mtw

<Normality Test>

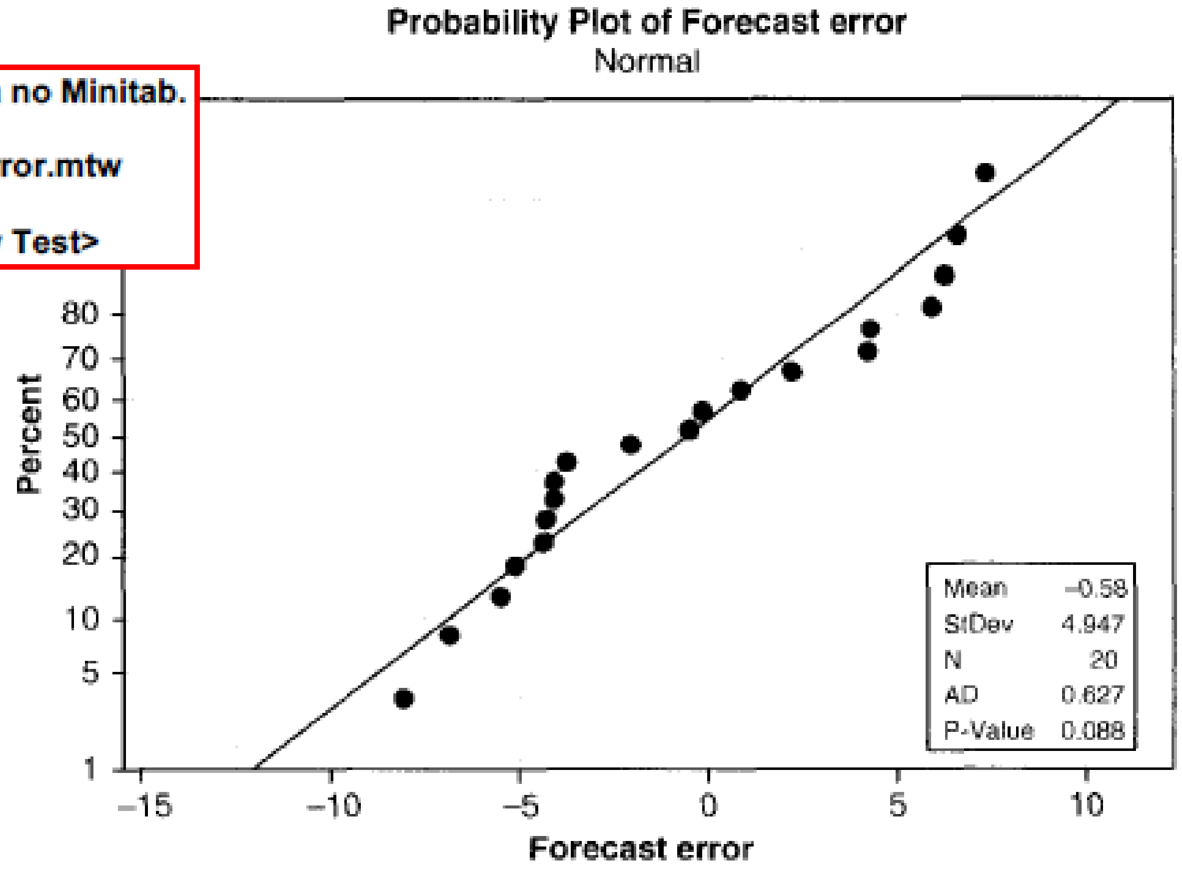


FIGURE 2.30 Normal probability plot of forecast errors from Table 2.2.

- If the sample ACF suggests that the forecast errors are not random, then this is evidence that the forecasts can be improved by refining the forecasting model.

- If a time series consists of uncorrelated observations and has constant variance, we say that it is **white noise.** If, in addition, the observations in this time series are normally distributed, the time series is **Gaussian white noise.**

- If a time series is white noise, the distribution of the sample autocorrelation coefficient at lag k in large samples is approximately normal with mean zero and variance $\frac{1}{T}$

$$r_k \sim \left(0, \frac{1}{T}\right)$$

- Therefore we could test the hypothesis $H_0: \rho_k = 0$ using the test statistic

$$Z_0 = r_k \sqrt{T}$$

This plot does not indicate any serious problem, with the normality assumption. so the forecast errors are Gaussian white noise.

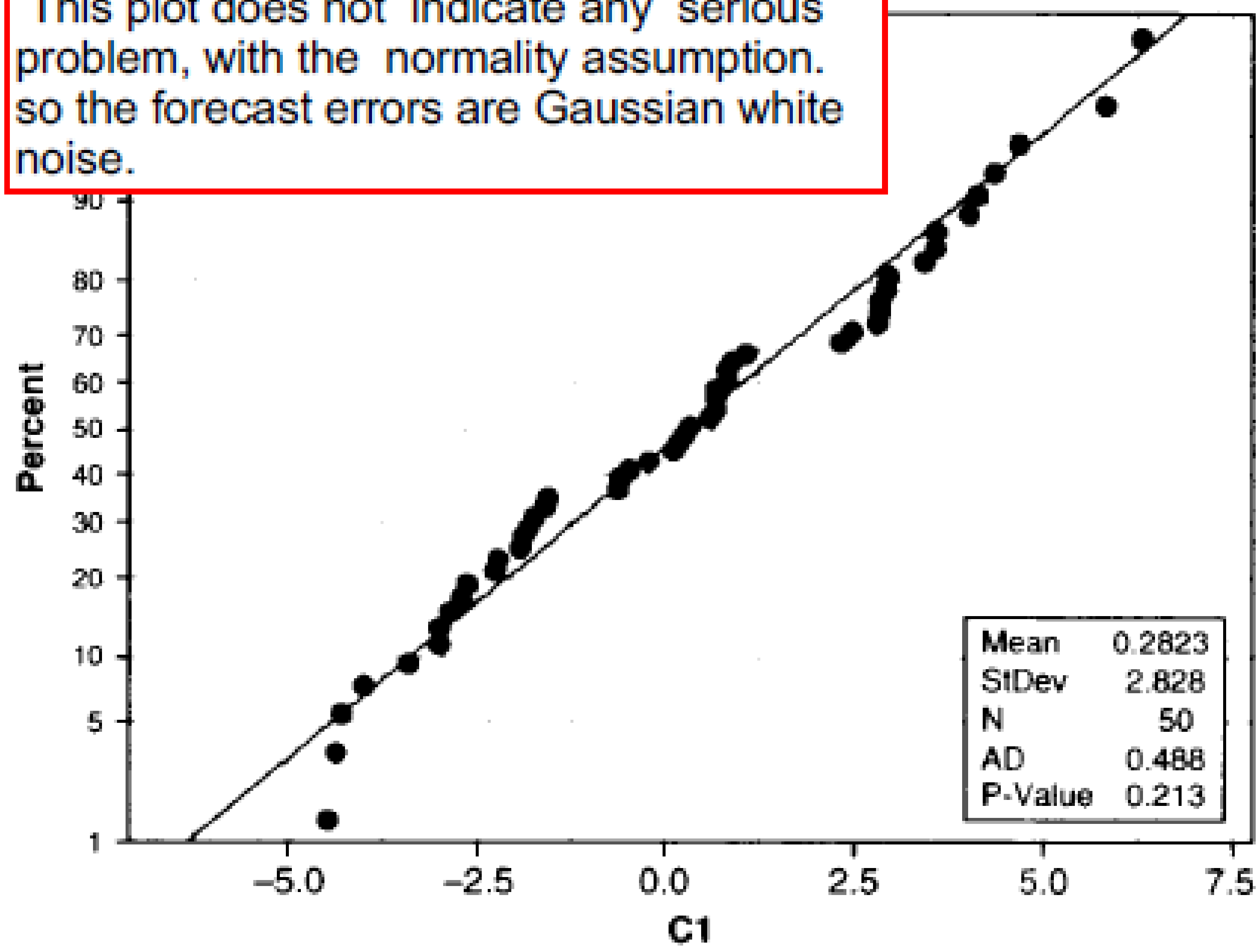


FIGURE 2.32 Normal probability plot of forecast errors from Table 2.3.

- We are often interested in evaluating a set of autocorrelations jointly to determine if they indicate that the time series is white noise.
- **Box and Pierce statistic**

$$Q_{BP} = T \sum_{k=1}^{K} r_k^2$$

is distributed approximately as chi-squared with K degrees of freedom under the null hypothesis that the time series is white noise. Therefore, if $Q_{BP} > \mathrm{X}^2_{\alpha,K}$ we would reject the null hypothesis and conclude that the time series is not white noise because some of the autocorrelations are not zero.

Goodness of fit statistic

- When the Box-Pierce statistic is applied on the set of residual autocorrelations, the statistic $Q_{BP} \sim \mathrm{X}^2_{\alpha, K-p}$, where p is the number of parameters in the model.

- Ljung-Box goodness of fit statistic (for small samples)

$$Q_{LB} = T(T+2) \sum_{k=1}^{K} \left(\frac{1}{T-k} \right) r_k^2$$

coefficient differs from zero. This procedure is a one-at-a-time test

BLE 2.4 Sample ACF of the One-Step-Ahead Forecast Errors in Table 2.3

Lag	Sample ACF, r_k	Z-Statistic	Ljung–Box Statistic, Q_{LB}
1	0.004656	0.03292	0.0012
2	−0.102647	−0.72581	0.5719
3	0.136810	0.95734	1.6073
4	−0.033988	−0.23359	1.6726
5	0.118876	0.81611	2.4891
6	0.181508	1.22982	4.4358
7	−0.039223	−0.25807	4.5288
8	−0.118989	−0.78185	5.4053
9	0.003400	0.02207	5.4061
10	0.034631	0.22482	5.4840
11	−0.151935	−0.98533	7.0230
12	−0.207710	−1.32163	9.9749

There is no strong evidence to indicate that the first 13 autocorrelations of the forecast errors considered **jointly** differ from zero, using Qlb.

Valor Critico =22,36

When evaluating the fit of the model to historical data, there are several criteria that may be of value.

+ mean squared error of the residuals

$$s^2 = \frac{\sum_{t=1}^{T} e_t^2}{T - p}$$

where T periods of data have been used to fit a model with p parameters and e_t is the residual from the model-fitting process in period t. An estimator of the variance of the model errors.

+ R-squared statistic

$$R^2 = 1 - \frac{\sum_{t=1}^{T} e_t^2}{\sum_{t=1}^{T}(y_t - \bar{y})^2}$$

Large values of R^2 suggest a good fit to the historical data.

- Adjusted R^2 statistic

$$R^2_{Adj} = 1 - \frac{\sum_{t=1}^{T} e_t^2/(t-p)}{\sum_{t=1}^{T}(y_t - \bar{y})^2/(T-1)}$$

+ Two other important criteria are the Akaike Information Criterion (AIC) (see Akaike [1974]) and the Schwarz Information Criterion (SIC) (see Schwarz [1978]):

$$AIC = \ln\left(\frac{\sum_{t=1}^{T} e_t^2}{T}\right) + \frac{2p}{T}$$

$$SIC = \ln\left(\frac{\sum_{t=1}^{T} e_t^2}{T}\right) + \frac{p \ln T}{T}$$

These two criteria penalize the sum of squared residuals for including additional parameters in the model. Models that have small values ofthe AIC or SIC are considered good models.

Monitoring a Forecasting Model

+ The simplest way is to apply **Shewhart control charts** to the forecast errors.
  - a plot of the forecast errors versus time containing a center line that represents the average (or the target value) of the forecast errors and a set of control limits that are designed to provide an indication that the forecasting model performance has changed.
  - The center line is usually taken as either zero or the average forecast error (ME)
  - The control limits are typically placed at three standard deviations of the forecast errors above and below the center line.
  - If the forecast errors plot within the control limits, we assume that the forecasting model performance is satisfactory (or in control), but if one or more forecast errors exceed the control limits, that is a signal that something has happened and the forecast errors are no longer fluctuating around zero.
  - In control chart terminology, we would say that the forecasting process is out of control and some analysis is required to determine what has happened.

- Moving range

$$|e_t(1) - e_{t-1}(1)|$$

- Moving range based on n observations

$$MR = \sum_{t=2}^{n} |e_t(1) - e_{t-1}(1)|$$

- The estimate of the standard deviation of the one step ahead forecast errors used to create the control limits.

$$\hat{\sigma}_{e(1)} = \frac{0.8865MR}{n-1}$$

Tere is no reason to suspect that the forecasting model is performing inadequately, at least from the statistical stability viewpoint. Forecast errors that plot outside the control limits would indicate model inadequacy, or possibly the presence of unusual observations such as outliers in the data. An investigation would be required to determine why these forecast errors exceed the control limits

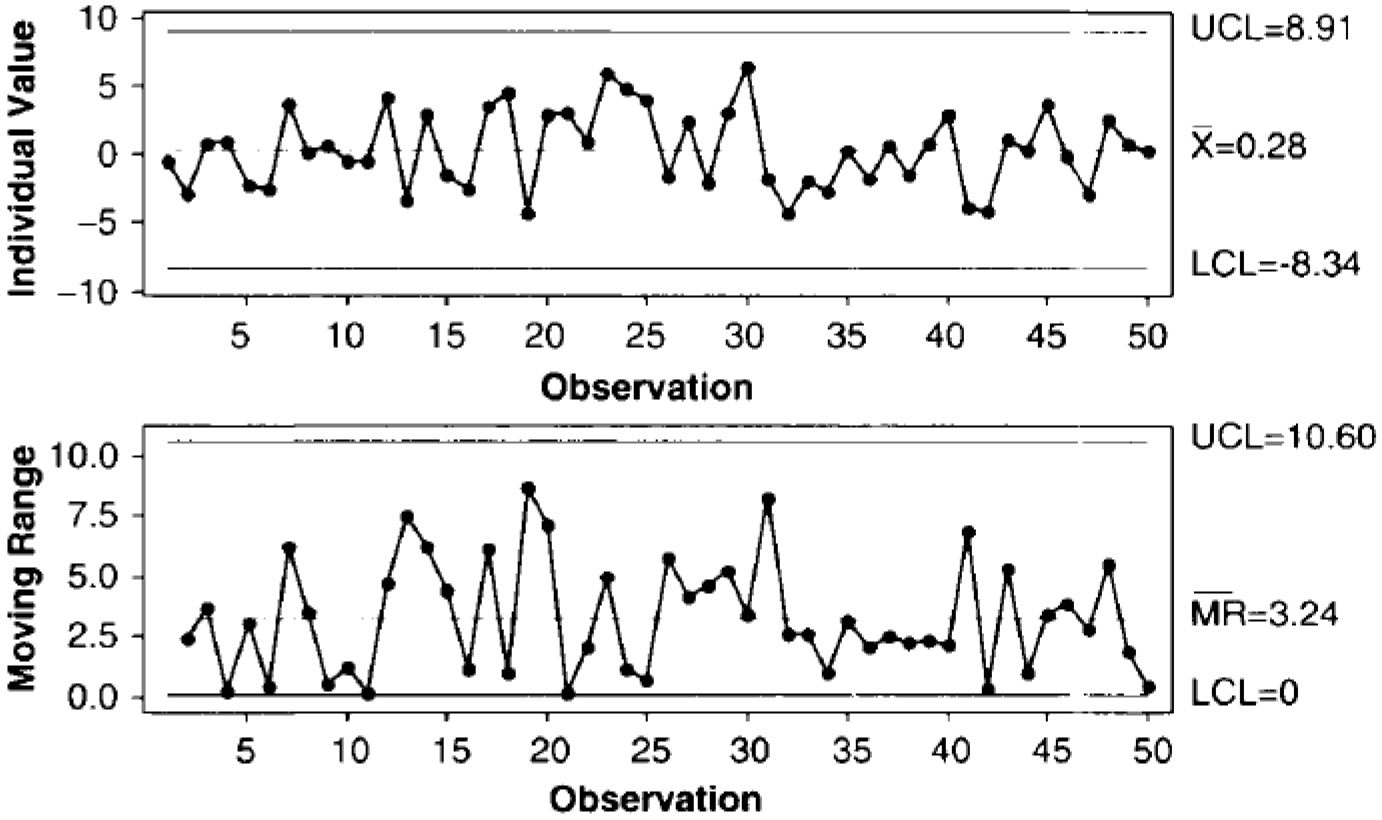


FIGURE 2.33 Individuals and moving range control charts of the one-step-ahead forecast errors in Table 2.3

- Two other types of control charts. the cumulative sum (or CUSUM) control chart and the exponentially weighted moving average (or EWMA) control chart. can also be useful for monitoring the performance of a forecasting model.
- These charts are more effective at detecting smaller changes or disturbances in the forecasting model performance than the individuals control chart.

- The CUSUM is very effective in detecting level.
- It works by accumulating deviations of the forecast errors that are above the desired target value T(usually either zero or the average forecast error), C^+upper CUSUM and deviations that are below the target, C^- lower CUSUM

$$C_t^+ = \max[0, e_t(1) - (T + K) + C_{t-1}^+]$$
$$C_t^- = \min[0, e_t(1) - (T + K) + C_{t-1}^+]$$

where the constant K is usually chosen as $K = 0.5\sigma_{e(1)}$

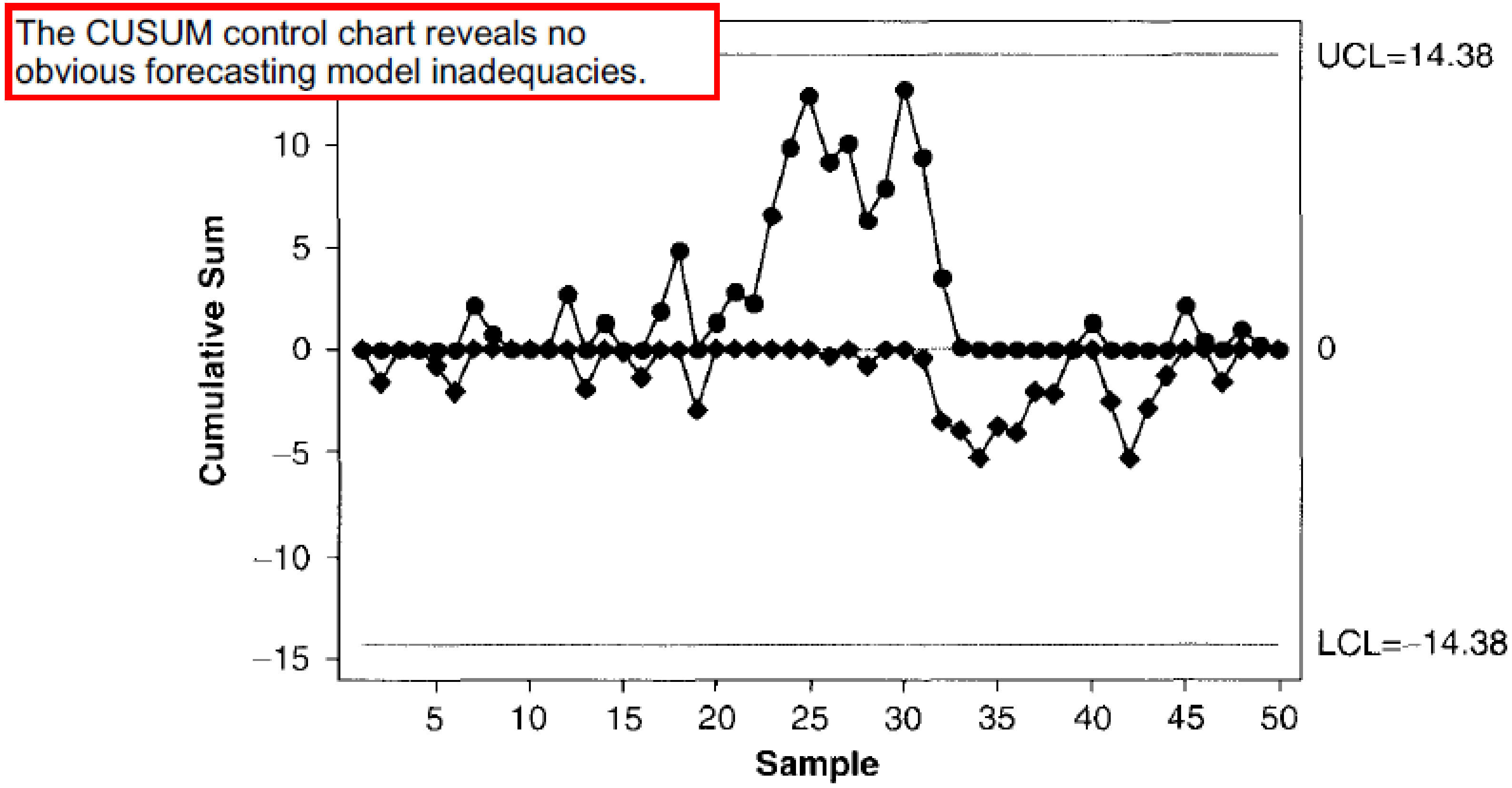


FIGURE 2.34 CUSUM control chart of the one-step-ahead forecast errors in Table 2.3.

EWMA

$$\bar{e}_t(1) = \lambda e_t(1) + (1-\lambda)\bar{e}_{t-1}(1)$$

where $\lambda > 0$ is the smoothing constant and the starting value of the EWMA (either $\bar{e}_0(1) = 0$ or the average of the forecast errors. Typically $0.05 < \lambda < 0.2$.

The EWMA is a weighted average of all current and previous forecast errors, and the weights decrease geometrically with the "age" of the forecast error.

$$\bar{e}_n(1) = \lambda \sum_{j=0}^{n-1} (1-\lambda)^j\, e_t(1) + (1-\lambda)^n \bar{e}_0(1)$$

Standard deviation of EWMA

$$\sigma_{\bar{e}_t(1)} = \sigma_{e(1)} \sqrt{\frac{\lambda}{2-\lambda}\left[1-(1-\lambda)^{2t}\right]}$$

EWMA control chart

$$UCL = T + 3\sigma_{\bar{e}_t(1)}$$
$$Center\ line = T$$
$$LCL = T - 3\sigma_{\bar{e}_t(1)}$$

None of the forecast errors exceeds the control limits so there is no indication of a problem with the forecasting model.

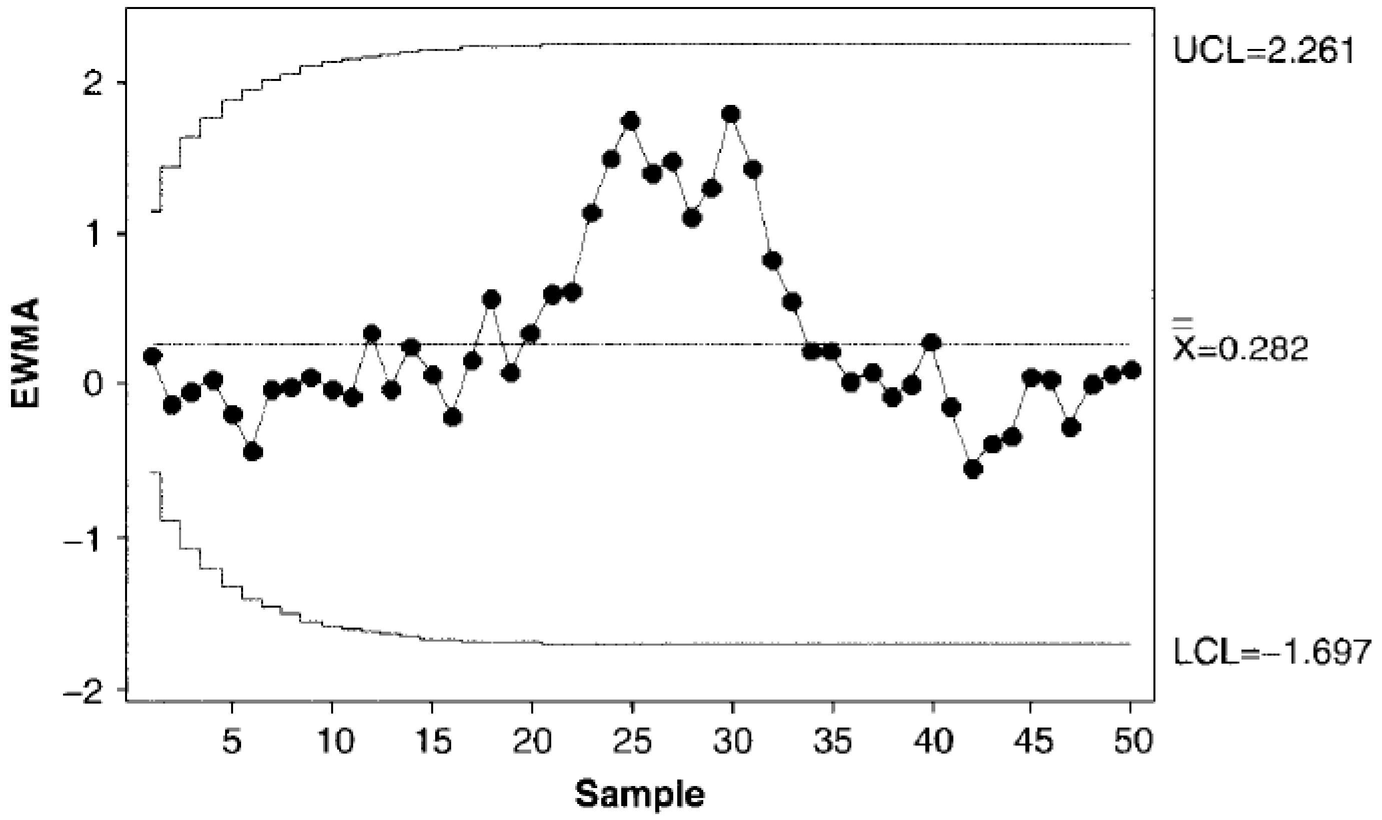


FIGURE 2.35 EWMA control chart of the one-step-ahead forecast errors in Table 2.3.

Tracking Signals

+ Cumulative sum of previous forecast errors

$$Y(n) = \sum_{t=1}^{n} e_t(1)$$

If the forecasts are unbiased we would expect $Y(n)$ to fluctuate around zero. If it differs from zero by very much, it could be an indication that the forecasts are biased.

+ Cumulative error tracking signal (CETS)

$$CETS = \left| \frac{Y(n)}{\hat{\sigma}_{Y(n)}} \right|$$

+ Smoothed error tracking signal

$$SETS = \left| \frac{\bar{e}_n(1)}{\hat{\sigma}_{\bar{e}_n(1)}} \right|$$

If the SETS exceeds a constant, say, K_2, this is an indication that the forecasts are biased and that there are potentially problems with the forecasting model.