Times Series and Forecasting (III)

Chapter 3. Modeling Deterministic Trends

Jianhua Hu

School of Statistics and Management
Shanghai University of Finance and Economics

Spring 2013

3.1. Trend and detrending

- Our goal is to consider the equally time space discrete stochastic processes $\{Y_t : t = 0, 1, \cdots, n\}$, time series models.

- Our interest is in realizations of stationary time series models.

- A time series process is stationary if the process satisfies
  - the mean function $\mu_t = \mathsf{E}(Y_t)$ is constant, and
  - $\gamma_{t,t-k}$ depends only on k, free of time t.

- However, many time series data sets "seem to" come from nonstationary time series processes.

- These data sets usually exhibit a trend associated with time t, or, "a long-term change with time t in the mean level".

Remark 1 on trend

- An obvious difficulty with the definition of a trend above is deciding what is meant by the phrase "long-term".
- For example, climatic processes can display cyclical variation over a long period of time, say, 1000 years. However, if one has just 30-40 years of data, this long-term cyclical pattern might be missed and interpreted as a trend which is linear.
- Also see the plot of the global temperature data.

Remark 2 on trend

- An analyst may mistakenly conjecture that a trend exists when it really does not.
- For example, in Figure 2.2, we have four realizations of a N(0,1) walk process

$$Y_t = Y_{t-1} + e_t.$$

  There is no trend in the mean of this random walk process. However, it would be easy to incorrectly assert that trends are present. See plots for the random walk.

- Random walks are non-stationary because $\gamma_{t,t-k}$ depends on both time t and time lag k.

Remark 3 on trend

- On the other hand, an analyst may mistakenly conjecture that a trend does not exist when it really does. In this case, the data are very noisy.

- For example, the lower right plot in Figure 2.6 is a noisy realization of a sinusoidal process considered in the last chapter. It is easy to miss the true cyclical structure from looking at the plot. See the plots in Figure 2.6.

Four examples introduced in Chapter 1

Figure 3.1 displays four time series we have seen from Chapter 1. Which (if any) type of trends do you think are present in each?

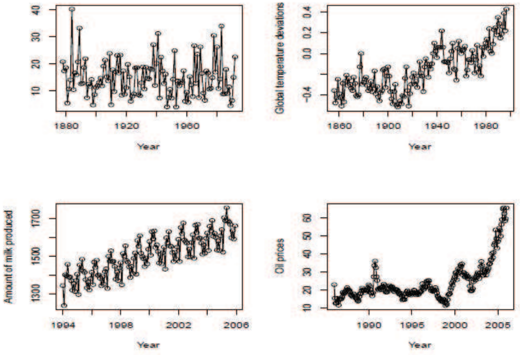


Figure 3.1: *Four time series. Upper left: LA rain data. Upper right: Global temperature data. Lower left: Milk production data. Lower right: Oil production data.*

Deterministic trend models

- Consider models of the form

$$Y_t = \mu_t + X_t = \text{a deterministic trend} + \text{random error}$$

where μ_t is a deterministic function that describes the trend and X_t is random error, not necessarily a white noise.

- So $\mathsf{E}(X_t) = 0$ for all t (a common assumption), and

$$\mathsf{E}(Y_t) = \mu_t$$

is the mean function for the process $\{Y_t\}$.

Two deterministic trend functions

- The deterministic trend is described by a kth order polynomial in time

$$\mu_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \cdots + \beta_k t^k.$$

The simplest two cases are linear and quadratic trends.

- If the trend is cyclical, consider functions of the form

$$\mu_t = \beta_0 + \sum_{j=1}^{m} \left(\alpha_j \cos \omega_j t + \beta_j \sin \omega_j t\right),$$

where the α_j's and β_j's are regression parameters and the ω_j's are related to frequencies of the trigonometric functions $\cos \omega_j t$ and $\sin \omega_j t$.

Two ways for removing trend or detrending

- Our goal is to deal with stationary time series models for data.
- Remove the trend if there is a deterministic trend present in the stochastic process.
- There are two general methods (I and II) to remove a trend.

Method I: using estimate to remove a trend

- Step 1. Estimate the trend and then subtract the estimated trend from the data (perhaps after transforming the data). Specifically, estimate μ_t with $\hat{\mu}_t$ and then model the residuals $\hat{X}_t = Y_t - \hat{\mu}_t$ as a stationary process (expected to be).
- Step 2. Implement standard diagnostics on the residuals $\hat{X}_t$ to check for violations of stationarity and other assumptions.
- Step 3. If the residuals are stationary, we can use a stationary time series model (Chapter 4) to describe their behavior.
- Step 4. Forecasting takes place by first forecasting the residual process $\hat{X}_t$ and then inverting the transformations described above to arrive back at forecasts for the original series $\{Y_t\}$.

Method II: differencing to a series for detrending

- Step 1. Apply repeatedly differencing to the series $\{Y_t\}$ until the differenced observations resemble a realization of some stationary time series.
- The approach was developed extensively to ARIMA models by Box and Jenkins (1976).
- Step 2. Use the theory of stationary processes for the modeling, analysis, and prediction of the stationary series and then transform this analysis back in terms of the original series $\{Y_t\}$ (studied in Chapter 5).

Remark for detrending

Important

If we assert that a trend exists and we fit a deterministic model that incorporates it, we are implicitly assuming that

the trend lasts "forever".

In some applications, this might be reasonable, but probably not in most.

3.2. A constant "trend" model

- Consider the model
$$Y_t = \mu + X_t,$$
where μ is constant (free of t) and where $\mathrm{E}(X_t) = 0$.
- The most common (LS) estimator of μ is
$$\overline{Y} = \frac{1}{n}\sum_{t=1}^{n} Y_t,$$
the sample mean, an unbiased estimator of μ.
- To assess the precision of $\overline{Y}$ as an estimator of μ, we examine the variance $\mathrm{var}(\overline{Y})$. If $\{Y_t\}$ is a stationary process with autocorrelation function (ACF) ρ_k's, then
$$\mathrm{var}(\overline{Y}) = \frac{\gamma_0}{n}\left[1 + 2\sum_{k=1}^{n-1}\left(1-\frac{k}{n}\right)\rho_k\right],$$ see Exercise 2.17
where $\mathrm{var}(Y_t) = \gamma_0$.

Variance of $\overline{Y}$ for μ and comparison to the iid case

- If $\{Y_t\}$ is an iid process, that is, $Y_1, Y_2, \cdots, Y_n$ is an iid (random) sample, then

$$\text{var}(\overline{Y}) = \gamma_0/n = \sigma_e^2/n.$$

- Therefore, var$(\overline{Y})$ in general can be larger than or smaller than γ_0/n depending on the values of ρ_k through

$$\frac{\gamma_0}{n}\left[1 + 2\sum_{k=1}^{n-1}\left(1 - \frac{k}{n}\right)\rho_k\right] - \frac{\gamma_0}{n} = \frac{2\gamma_0}{n}\sum_{k=1}^{n-1}\left(1 - \frac{k}{n}\right)\rho_k.$$

  - If this quantity is smaller than zero, i.e., var$(\overline{Y}) < \gamma_0/n$, $\overline{Y}$ is a better estimator of μ than $\overline{Y}$ is in an iid sampling context.
  - If this quantity is larger than zero, i.e., var$(\overline{Y}) > \gamma_0/n$, $\overline{Y}$ is a worse estimator of μ than $\overline{Y}$ is in an iid sampling context.

Example 3.1. A constant trend of a MA process

- Suppose that $\{Y_t\}$ is a MA process given by

$$Y_t = \frac{1}{3}(e_t + e_{t-1} + e_{t-2}),$$

where $\{e_t\} \sim WN(0, \sigma_e^2)$.

- In the last chapter, we calculated

$$\gamma_k = \begin{cases} \sigma_e^2/3, & k = 0 \\ 2\sigma_e^2/9, & k = 1 \\ \sigma_e^2/9, & k = 2 \\ 0, & k > 2. \end{cases}$$

- The lag 1 autocorrelation for this process is

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{2\sigma_e^2/9}{\sigma_e^2/3} = 2/3.$$

Example 3.1. (Continued)

- The lag 2 autocorrelation for this process is

$$\rho_2 = \frac{\gamma_2}{\gamma_0} = \frac{\sigma_e^2/9}{\sigma_e^2/3} = 1/3.$$

  Also, $\rho_k = 0$ for all $k > 2$.

- Substituting these values in and simplifying, we get

$$\begin{aligned} \text{var}(\overline{Y}) &= \frac{\gamma_0}{n}\left[1 + 2\sum_{k=1}^{n-1}\left(1 - \frac{k}{n}\right)\rho_k\right] \\ &= \frac{\gamma_0}{n} + \frac{4(n-1)\gamma_0 + 2(n-2)\gamma_0}{3n^2} > \frac{\gamma_0}{n}. \end{aligned}$$

- Therefore, we lose efficiency in estimating μ with $\overline{Y}$ when compared to using $\overline{Y}$ in an iid sampling context. The positive ACF make estimation of μ more inefficiency.

Example 3.2. Constant trend of an AR model

- Suppose that $\{Y_t\}$ is a stationary process with ACF $\rho_k = \phi^k$, where $-1 < \phi < 1$.
  - The ACF decays exponentially as the lag k increases.
  - AR(1) possesses this ACF. We will see that in Chapter 4.
- To examine the effect of estimating μ with $\overline{Y}$ in this situation, we use an approximation for var$(\overline{Y})$ for large n, specifically,

$$\text{var}(\overline{Y}) = \frac{\gamma_0}{n}\left[1 + 2\sum_{k=1}^{n-1}\left(1 - \frac{k}{n}\right)\rho_k\right] \approx \frac{\gamma_0}{n}\left[1 + 2\sum_{k=1}^{\infty}\rho_k\right],$$

where we have taken $(1 - k/n) \approx 1$ for n large.

Example 3.2. (Continued)

- Therefore,

$$\text{var}(\overline{Y}) \approx \frac{\gamma_0}{n}\left[1 + 2\sum_{k=1}^{\infty}\rho_k\right] = \frac{\gamma_0}{n}\left[1 + 2\left(\sum_{k=0}^{\infty}\phi^k - 1\right)\right]$$
$$= \frac{\gamma_0}{n}\left[1 + 2\left(\frac{1}{1-\phi} - 1\right)\right] = \left(\frac{1+\phi}{1-\phi}\right)\frac{\gamma_0}{n}.$$

- For example, if $\phi = -0.6$, then

$$\text{var}(\overline{Y}) \approx 0.25\frac{\gamma_0}{n} < \gamma_0/n.$$

- For this process, using $\overline{Y}$ produces a better estimate of μ than in the iid sampling situation.

Example 3.3. A constant trend of a random walk

- Consider the random walk process

$$Y_t = Y_{t-1} + e_t.$$

- This process is not stationary. So we can not use the var$(\overline{Y})$ formula presented earlier.

- However, recall that this process can be written out as

$$Y_n = e_1 + e_2 + \cdots + e_n, \text{ for } n = 1, 2, \cdots$$

so that

$$\overline{Y} = \frac{1}{n}\sum_{t=1}^{n} Y_t = \frac{1}{n}\sum_{t=1}^{n}(n+1-t)e_t.$$

Example 3.3 (Continued)

Therefore,

$$
\begin{aligned}
\mathsf{var}(\overline{Y}) &= \frac{1}{n^2}\sum_{t=1}^{n}(n+1-t)^2\mathsf{var}(e_t) \\
&= \frac{\sigma_e^2}{n^2}\left[1^2+2^2+\cdots+(n-1)^2+n^2\right] \\
&= \frac{\sigma_e^2}{n^2}\left[\frac{n(n+1)(2n+1)}{6}\right] \\
&= \frac{\sigma_e^2}{n}\left[\frac{(n+1)(2n+1)}{6}\right].
\end{aligned}
$$

Remarks on Example 3.3

- This result is surprising! Note that as n increases, so does var$(\overline{Y})$. That is, averaging a larger sample produces a worse (i.e., more variable) estimate of μ than averaging a smaller one!!
- This is quite different than the results obtained for stationary processes. The nonstationarity in the data causes very bad things to happen, even in the relatively simple task of estimating an overall process mean.

CI and hypothesis test for μ

- Suppose that $Y_t = \mu + X_t$ with $\{X_t\} \sim N(0, \gamma_0)$.
- Thus, $Y_t \sim N(\mu, \gamma_0)$ and $\{Y_t\}$ is stationary. Therefore,

$$\overline{Y} \sim N\left\{\mu, \frac{\gamma_0}{n}\left[1 + 2\sum_{k=1}^{n-1}\left(1 - \frac{k}{n}\right)\rho_k\right]\right\}.$$

- If γ_0 and the ρ_k's are known, then a $100(1-\alpha)$ percent confidence interval for μ is

$$\overline{Y} \pm z_{\alpha/2}\sqrt{\frac{\gamma_0}{n}\left[1 + 2\sum_{k=1}^{n-1}\left(1 - \frac{k}{n}\right)\rho_k\right]}.$$

- Task of hypothesis testing about μ can be done.

CI and hypothesis test for μ(Continued)

- The impact of the ACF ρ_k will be the same on CI as on estimate.
- More negative ACF ρ_k will make the standard error

$$\text{se}(\overline{Y}) = \sqrt{\frac{\gamma_0}{n}\left[1 + 2\sum_{k=1}^{n-1}\left(1 - \frac{k}{n}\right)\rho_k\right]}$$

  smaller, which will make the CI more precise (i.e., shorter).
- More positive ACF will make this quantity larger, thereby lengthening CI, making them less informative.

Remark on CI of the mean

- If $\rho_k = 0$, for all k, then this CI formula reduces to the iid case with
$$\overline{Y} \pm z_{\alpha/2}\sqrt{\frac{\gamma_0}{n}}.$$

- With real data, γ_0 and the ρ_k's will rarely be known. This implies that confidence interval is difficult to be obtained.

- We will talk about estimation of γ_0 and the autocorrelations ρ_k both in large and finite samples later.

3.3.1. Straight line model

- Consider the deterministic time trend model ($\mathsf{E}(X_t) = 0$)

$$Y_t = \mu_t + X_t = \beta_0 + \beta_1 t + X_t.$$

- One of the main reasons for fitting the straight line model was to capture the linear trend.
- Use a simple linear regression for the $\{Y_t\}$, where t is the predictor.
- "Fitting this model" means to estimate the regression parameters β_0 and β_1 using the observed data $\{y_t\}$. The X_t's are random errors and not observed.

LS method for straight line model

- Use the method of least squares, we get the least squares estimators of β_0 and β_1 as follows

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = (X'X)^{-1}X'Y$$

where $X = \begin{pmatrix} 1 & \cdots & 1 \\ t_1 & \cdots & t_n \end{pmatrix}'$ and $Y = \begin{pmatrix} Y_1 & \cdots & Y_n \end{pmatrix}'$.

- Specifically,

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{t} \text{ and } \hat{\beta}_1 = \frac{\sum_{t=1}^{n} (t - \overline{t}) Y_t}{\sum_{t=1}^{n} (t - \overline{t})^2}.$$

LS method for straight line model (Continued)

- If $\mathsf{E}(X_t) = 0$, $\{X_t\}$ independent, and $\mathsf{var}(X_t) = \gamma_0$ (a G-M model), by G-M Theorem, $\hat{\boldsymbol{\beta}} = (X'X)^{-1}XY$ is BLUE to $\boldsymbol{\beta}$ with $\mathsf{Cov}(\hat{\boldsymbol{\beta}}) = \gamma_0(X'X)^{-1}$. Specifically,

$$\mathsf{var}(\hat{\beta}_0) = \gamma_0 \left(\frac{1}{n} + \frac{\bar{t}^2}{\sum_{t=1}^{n}(t-\bar{t})^2} \right), \mathsf{var}(\hat{\beta}_1) = \frac{\gamma_0}{\sum_{t=1}^{n}(t-\bar{t})^2}.$$

- Note that a zero mean white noise process $\{X_t\}$ satisfies these assumptions.

LS method for straight line model (Continued)

- If $\mathrm{E}(X_t) = 0$, $\{X_t\}$ independent, $\mathrm{var}(X_t) = \gamma_0$, and the X_t's are normally distributed (a normal G-M model), then

$$\hat{\boldsymbol{\beta}} = (X'X)^{-1}XY \text{ is MVUE to } \boldsymbol{\beta}$$

with $\hat{\boldsymbol{\beta}} \sim N(\mathbf{0}, \gamma_0(X'X)^{-1})$, specifically,

$$\hat{\beta}_0 \sim N\left\{\beta_0, \gamma_0\left[\frac{1}{n} + \frac{\bar{t}^2}{\sum_{t=1}^{n}(t-\bar{t})^2}\right]\right\}$$

and

$$\hat{\beta}_1 \sim N\left\{\beta_1, \frac{\gamma_0}{\sum_{t=1}^{n}(t-\bar{t})^2}\right\}.$$

LS method for straight line model (Continued)

- On the errors X_t, zero mean, independence, homoscedasticity, and normality, are the usual assumptions on the errors in a standard regression setting!
- With most time series data sets, at least one of these assumptions will be violated, implying that some quantities provided in computing packages (e.g., SAS, R) may not be meaningful.
- The only instance in which these quantities are exactly correct is if $\{X_t\}$ is an iid normal white noise process.

Example 3.4. LS method for straight line model

- Consider the global temperature data in Figure 3.2. There is an apparent upward trend in the temperature data series, which has been used to argue the global warming hypothesis, at least over this time period.
- Estimate this trend by fitting the straight line regression model

$$Y_t = \beta_0 + \beta_1 t + X_t,$$

for $t = 1900, 1901, \cdots, 1997$.

- As usual, assume that $\mathrm{E}(X_t) = 0$. We can perform this analysis in SAS and do so now.

Example 3.4. SAS output from Example 1-1-1.sas

Analysis of Variance

Source	DF	SS	MS	F-value	Pr > F
Model	1	3.02386	3.02386	179.49	$<.0001$
Error	96	1.61730	0.01685		
Corrected T	97	4.64116			

Root MSE	0.12980	R-Square	0.6515
Dependent Mean	-0.08724	Adj R-Sq	0.6479
Coeff Var	-148.77132		

Parameter Estimates

Variable	DF	Estimate	SE	t-value	Pr > $\|t\|$
Intercept	1	-12.18641	0.90319	-13.49	$<.0001$
t (slope)	1	0.00621	0.00046348	13.40	$<.0001$

Figure 3.2

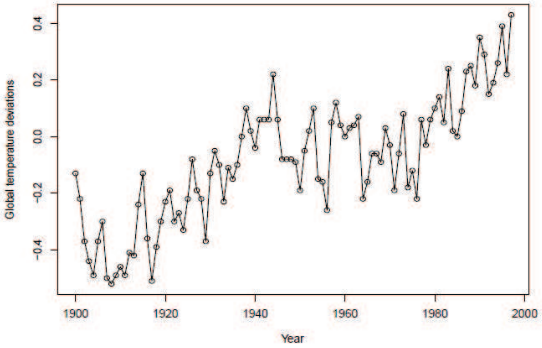


Figure 3.2: *Global temperature data from 1900-1997. The data are a combination of land-air average temperature anomalies, measured in degrees Centigrade.*

Analysis of SAS output

The least squares estimates are $\hat{\beta}_0 = -12.19$ and $\hat{\beta}_1 = 0.0062$ so that the fitted regression model is

$$\hat{Y}_t = -12.19 + 0.0062t.$$

This is the equation of the straight line superimposed over the series in Figure 3.3.

Figure 3.3

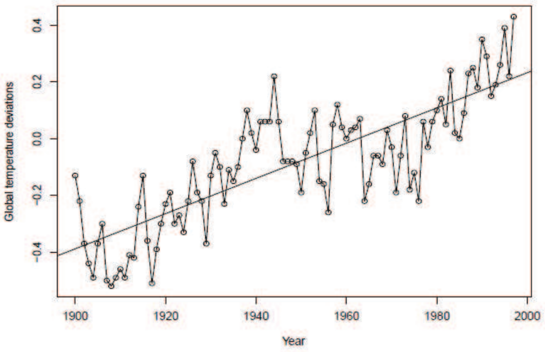


Figure 3.3: *Global temperature data (1900-1997) with a straight line trend fit. The data are a combination of land-air average temperature anomalies, measured in degrees C.*

Residuals from the least squares fit

- The residuals from the least squares fit are given by

$$\hat{X}_t = Y_t - \hat{Y}_t,$$

- In this example (with the straight line model fit), the residuals are given by

$$\hat{X}_t = Y_t - \hat{Y}_t = Y_t + 12.19 - 0.0062t,$$

for $t = 1900, 1901, \cdots, 1997.$

- The residual process $\hat{X}_t$ contains information in the data that is not accounted for in the straight line trend model. For this reason, it is called the detrended series.

Figure 3.4, Residuals from a simple regression line

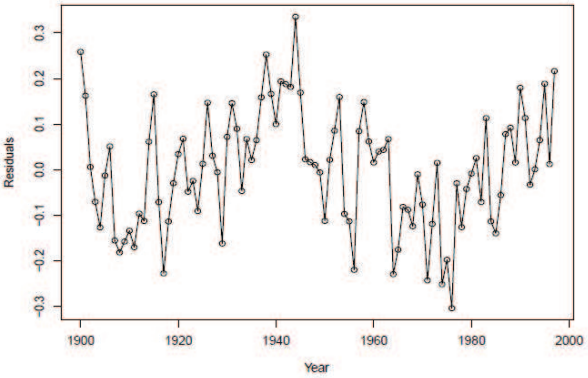


Figure 3.4: *Residuals from the straight line trend model for the global temperature data.*

Is residuals from the least squares fit stationary?

- From Figure 3.4, this detrended series does appear to be somewhat stationary, at least much more so than the original series $\{Y_t\}$.
- From the plot, the residuals are not white noise because a WN is stationary.
- Durbin-Watson test (1951, close to 2 for errors) for autocorrelation in time series data in the proc autoreg procedure can be used to test the global temperature data. SAS command is

  ```
  Proc autoreg data=; model reponse=t/dwprob; run;
  ```

  See figure3-4.sas for output for autocorrelation.

Example 3.4. Differencing method for strait linear model

- Examine the first difference process $\{\nabla Y_t = Y_t - Y_{t-1}\}$.
- Doing so here, as evidenced in Figure 3.5, produces a new process that does appear to be somewhat stationary (more so than the detrended residual process from fitting the straight line model).
- See sas program figure3-5.sas:

```
Proc autoreg data=; model residualD=t/dwprob; run;
```

Figure 3.5. Difference plot of global temperature data

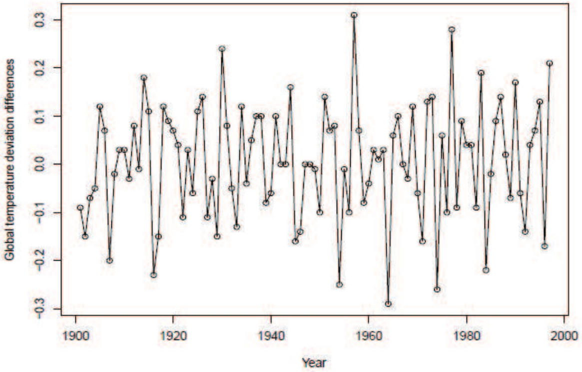
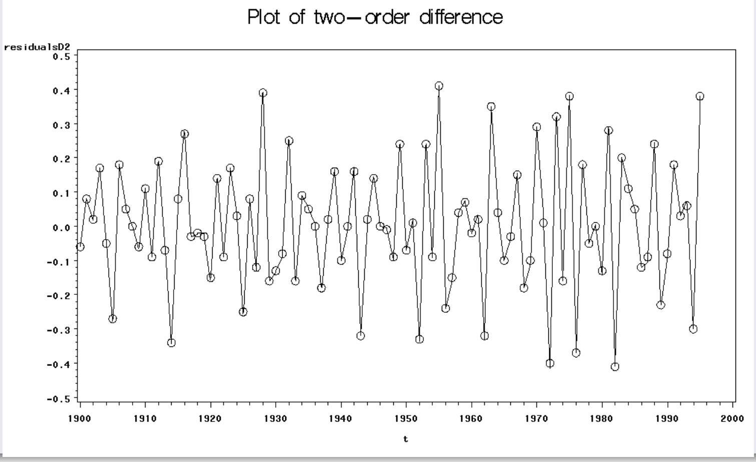


Figure 3.5: *Global temperature data differences (1900-1997).*

Is higher order differencing needed for temperature data?

- It is not necessary for the second order differencing for the global temperature data.
  - Check quadratic trend for the data by using proc reg.
  - Or examine the second difference process $\{\nabla^2 Y_t = Y_t - 2Y_{t-1} + Y_{t-2}\}$ by using proc reg.
- Why? See Exercise 2.9:
  If $Y_t = \beta_0 + \beta_1 t + X_t$ with $\mathsf{E}(X_t) = 0$ is non-stationary,
  then $\Delta Y_t = Y_t - Y_{t-1}$ is stationary.
  So is $\Delta^k Y_t = \Delta^{k-1} Y_t - \Delta^{k-1} Y_{t-1}$ for $k > 1$.

Figure 3.5-1. The second-order difference plot of the data



Advantage and disadvantage of two approaches

Both detrending and differencing can be helpful in transforming a non-stationary process into (or at least appears) stationary one.

- One advantage of differencing over detrending to remove trend is that no parameters are estimated in taking differences.
- One disadvantage of differencing is that it does not provide an "estimate" of the error process X_t.
- If an estimate of the error process is crucial, detrending may be more appropriate. If the goal is only to coerce the data to stationarity, differencing may be preferred.

3.3.2. Polynomial regression

- Consider the deterministic time trend model ($\mathsf{E}(X_t) = 0$)

$$Y_t = \mu_t + X_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \cdots + \beta_k t^k + X_t.$$

- The μ_t is a polynomial function with degree $k \geq 1$.
  - If $k = 1$, $\mu_t = \beta_0 + \beta_1 t$ is a linear trend function.
  - If $k = 2$, $\mu_t = \beta_0 + \beta_1 t + \beta_2 t^2$ is a quadratic trend function.
  - If $k = 3$, $\mu_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3$ is a cubic trend function,
  - and so on.

LS estimation

Use the method of least squares, we get the least squares estimators of $\boldsymbol{\beta}$ as follows

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\beta}_0 & \hat{\beta}_1 & \cdots & \hat{\beta}_k \end{pmatrix}' = (X'X)^{-1}X'Y$$

where $X = \begin{pmatrix} 1 & t_1 & \cdots & t_1^k \\ \cdots & \cdots & \cdots & \cdots \\ 1 & t_n & \cdots & t_n^k \end{pmatrix}$ and $Y = \begin{pmatrix} Y_1 & \cdots & Y_n \end{pmatrix}'$.

- If $\mathrm{E}(X_t) = 0$, the LS estimator $\hat{\boldsymbol{\beta}}$ is unbiased to $\boldsymbol{\beta}$.
- If $X_t \sim (0, \sigma^2)$, $\hat{\boldsymbol{\beta}}$ is BLUE with $\mathrm{Cov}(\hat{\boldsymbol{\beta}}) = \sigma^2(X'X)^{-1}$.
- If $X_t \sim N(0, \sigma^2)$, $\hat{\boldsymbol{\beta}}$ is MVUE with $\mathrm{Cov}(\hat{\boldsymbol{\beta}}) = \sigma^2(X'X)^{-1}$.

Example 3.5. Gold price data in USD during the year 2005

- Observe a time series of $n = 254$ daily observations on the price of gold (per troy ounce) in US dollars during the year 2005.
- Data is from gold.dat in file package "Data_example".
- Look at the plot of gold price data. See Figure 3.6.

Figure 3.6. Plot of price of gold in USD during 2005

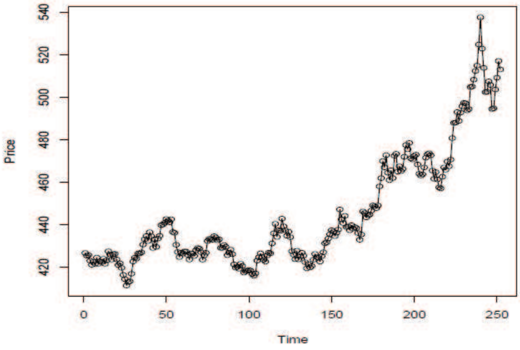


Figure 3.6: *Gold price data. Daily price in US dollars per troy ounce: 1/4/05 through 12/30/05.*

SAS output of Example 3.5 from Example3-5.sas

- From Figure 3.6, we will use a quadratic regression model

$$Y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + X_t,$$

for $t = 1, 2, \cdots, 254$, to fit the gold price data. Assume that $\mathsf{E}(X_t) = 0$.

- Use SAS program to detrend the data.

SAS output of Example 3.5 from Example3-5.sas

Analysis of Variance

Source	DF	SS	MS	F-value	Pr > F
Model	2	163655	81827	946.60	$< .0001$
Error	249	21525	86.44390		
Corrected T	251	185179			

Root MSE	9.29752	R-Square	0.8836
Dependent Mean	444.98790	Adj R-Sq	0.8828
Coeff Var	2.08939		

Parameter Estimates

Variable	DF	Estimate	SE	t-value	Pr > $\|t\|$
Intercept	1	434.60058	1.77110	245.38	$< .0001$
t (slope)	1	-0.36184	0.03233	-11.19	$< .0001$
t^2(curve)	1	0.00264	0.00012374	21.31	$< .0001$

The quadratic fit from SAS output for gold price data

- The fitted quadratic regression curve is

$$\hat{Y}_t = 434.6 - 0.362t + 0.00264t^2,$$

for $t = 1, 2, \cdots, 254$. The quadratic fit plot is depicted in Figure 3.7.

- The residual process is

$$\hat{X}_t = Y_t - \hat{Y}_t = Y_t - 434.6 + 0.362t - 0.00264t^2,$$

for $t = 1, 2, \cdots, 254$. The residual process is depicted in Figure 3.8.

Figure 3.7. The quadratic fitted regression of gold price

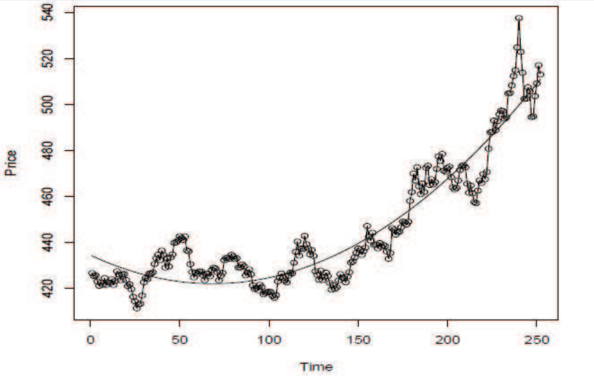


Figure 3.7: *Gold price data with a quadratic trend fit.*

Figure 3.8. Residuals from the quadratic trends fit

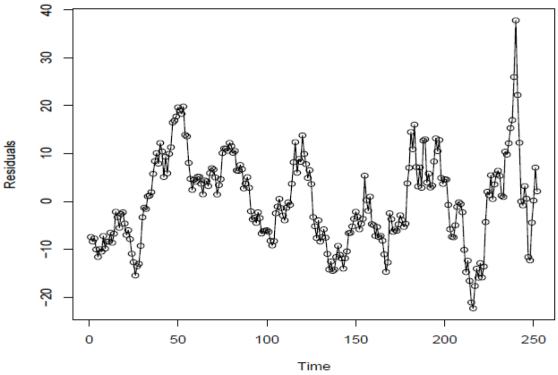


Figure 3.8: *Gold price data. Residuals from the quadratic trend fit.*

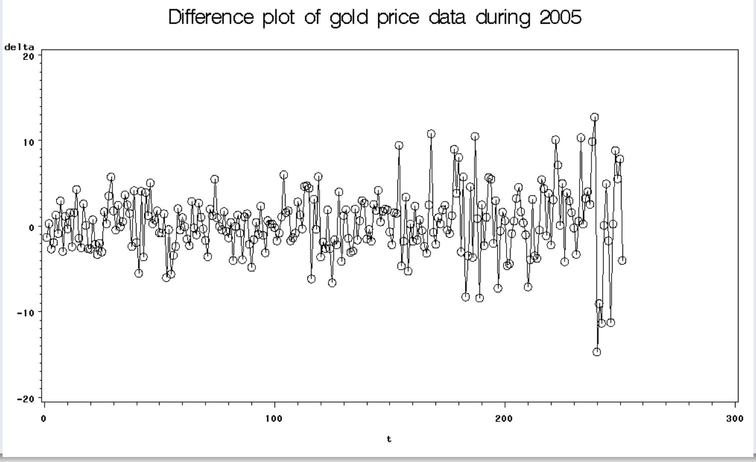
Analysis from the detrended (residual) series

- This detrended process from the detrended (residual) series appears to be somewhat stationary, much more so than the original process from the original time series.
- It should be obvious that the detrended (residual) process is not white noise. There is still a large amount of momentum left in the residuals. In fact, the autocorrelations are probably quite large at low lags.
- Use proc autoreg to check autocorrelation of the data. See output from sas program.

Example 3.5. Differencing method for gold price data

- Examine the first-order difference process $\{\nabla Y_t = Y_t - Y_{t-1}\}$.
- Doing so here produces a new process that does appear to be somewhat stationary (more so than the detrended residual process from fitting the quadratic model). Look at that evidenced in Figure 3.8-1.
- Is it necessary to need two-order difference process for the gold price data?

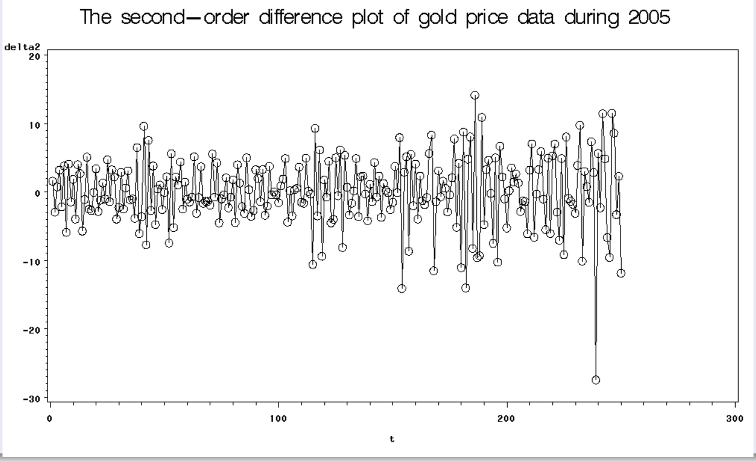
Figure 3.8-1. Difference plot of gold price data



Example 3.5. Is 2nd-order difference needed for gold price data

- Is it necessary to need two-order difference process for the gold price data?
- My answer is "No".
- Why?

Figure 3.8-2. The second difference plot of the data



3.3.3. Seasonal means model

- Consider the deterministic trend model

$$Y_t = \mu_t + X_t,$$

where $\mathsf{E}(X_t) = 0$ and where

$$\mu_t = \begin{cases} \beta_1, & t = 1, 13, 25, \cdots \\ \beta_2, & t = 2, 14, 26, \cdots \\ \vdots & \\ \beta_12, & t = 12, 24, 36, \cdots \end{cases}$$

- The regression parameters $\beta_1, \beta_2, \cdots, \beta_{12}$ are fixed constants. This is called the seasonal means model.

Seasonal means model (Continued)

- This model does not take the shape of the seasonal trend into account; instead, it merely says that observations 12 months apart ha ve the same mean, and this mean does not change through time.

- Other seasonal means models with a different number of parameters could be specified. For instance, for quarterly data, we could use a mean function with 4 regression parameters β_1, β_2, β_3 and β_4.

Fitting seasonal means model

Use the method of least squares, we get the least squares estimators of $\boldsymbol{\beta}$ as follows

$$\hat{\boldsymbol{\beta}} = (\hat{\beta}_0 \quad \hat{\beta}_1 \quad \cdots \quad \hat{\beta}_k)' = (X'X)^{-1}X'Y$$

where $X = (I_{12} \quad \cdots \quad I_{12})'$ (assume $n = 12n_1$) and $Y = (Y_1 \quad \cdots \quad Y_n)'$.

- If $\mathrm{E}(X_t) = 0$, the LS estimator $\hat{\boldsymbol{\beta}}$ is unbiased to $\boldsymbol{\beta}$.
- If $X_t \sim (0, \sigma^2)$, $\hat{\boldsymbol{\beta}}$ is BLUE with $\mathrm{Cov}(\hat{\boldsymbol{\beta}}) = \sigma^2(X'X)^{-1}$.
- If $X_t \sim N(0, \sigma^2)$, $\hat{\boldsymbol{\beta}}$ is MVUE with $\mathrm{Cov}(\hat{\boldsymbol{\beta}}) = \sigma^2(X'X)^{-1}$.

Fitting seasonal means model (Cont.)

- In particular,

$$\hat{\beta}_1 = \frac{1}{n_1} \sum_{t \in \mathbb{A}_1} Y_t,$$

  where the set $\mathbb{A}_1 = \{t : t = 1 + 12j, j = 0, 1, 2, \cdots, \}$. n_1 is the number of observations in month 1 (e.g., January).

- In general,

$$\hat{\beta}_i = \frac{1}{n_i} \sum_{t \in \mathbb{A}_i} Y_t,$$

  where the set $\mathbb{A}_i = \{t : t = i + 12j, j = 0, 1, 2, \cdots, \}$, for $i = 1, 2, \cdots, 12$, where n_i is the number of observations in month i.

Figure 3.9. Plot of monthly US beer sales data 1980-1990

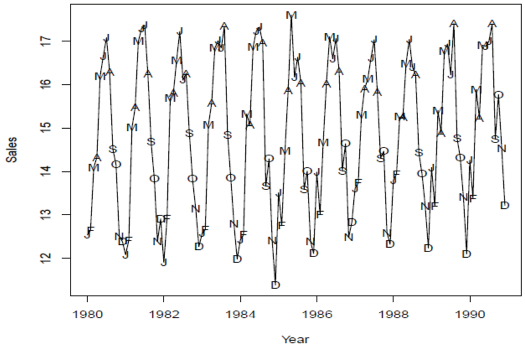


Figure 3.9: *Monthly US beer sales from 1980-1990. The data are measured in millions of barrels.*

Example 3.6. Monthly US beer sales data 1980-1990

The data in Figure 3.9 are monthly US beer sales (in millions of barrels) in United States during the period from January of 1980 to December of 1990.

Parameter estimate from Proc glm procedure

This time series has a relatively constant mean overall with the repeating patterns over time, so a seasonal means model may be appropriate. Fitting the model vis Proc glm procedure causes the results.

Parameter	Estimate	Standard Error	t-Value	Pr > |t|
January	13.1608091	0.16471777	79.90	<.0001
February	13.0175818	0.16471777	79.03	<.0001
March	15.1058182	0.16471777	91.71	<.0001
April	15.3981273	0.16471777	93.48	<.0001
May	16.7695273	0.16471777	101.81	<.0001
June	16.8791818	0.16471777	102.47	<.0001
July	16.8270091	0.16471777	102.16	<.0001
August	16.5716182	0.16471777	100.61	<.0001
September	14.4044545	0.16471777	87.45	<.0001
October	14.2847545	0.16471777	86.72	<.0001
November.	12.8943091	0.16471777	78.28	<.0001
December.	12.3403818	0.16471777	74.92	<.0001

Discussion for SAS output

- The only quantities that have relevance are the least squares estimates. The estimate $\hat{\beta}_i$ is simply the sample mean of the observations for month i; thus, $\hat{\beta}_i$ is an unbiased estimate of the ith (population) mean monthly sales β_i.
- A plot of the residuals from the seasonal means model fit is given in Figure 3.10. This residual process looks possibly stationary (I can detect a slight increasing trend).
- See SAS programming Example 3-6-0.sas

Figure 3.10. Residual from the seasonal mean model fit

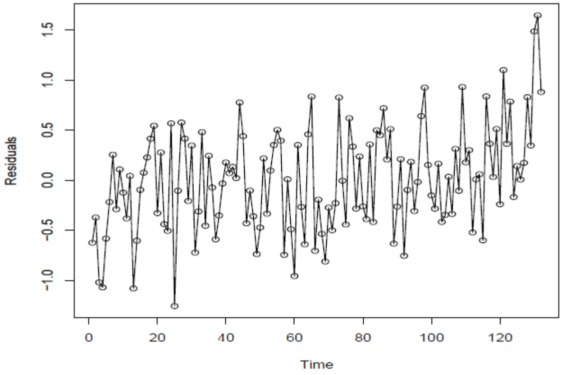


Figure 3.10: *Beer sales data. Residuals from the seasonal means model fit.*

3.3.4. Cosine trend model

- Consider the deterministic time trend model ($\mathsf{E}(X_t) = 0$)

$$Y_t = \mu_t + X_t = \beta \cos(2\pi f t + \Phi) + X_t.$$

- The trigonometric mean μ_t consists of different parts:
  - β is the amplitude. μ_t oscillates between $-\beta$ and β.
  - f is the frequency $\Longrightarrow 1/f$ is the period (the time it takes to complete one full cycle of the function). For monthly data, the period is 12 months; i.e., the frequency is $f = 1/12$.
  - Φ is the phase shift. This represents a horizontal shift in μ_t.

Transform cosine trend into a linear trend

- Let $\beta_1 = \beta\cos\Phi$ and $\beta_2 = -\beta\sin\Phi$, so that the phase shift

  $$\Phi = \tan^{-1}\left(-\frac{\beta_2}{\beta_1}\right)$$

  and the amplitude $\beta = \sqrt{\beta_1^2 + \beta_2^2}$. The rewritten expression,

  $$\mu_t = \beta_1\cos(2\pi ft) + \beta_2\sin(2\pi ft),$$

  is a linear function of β_1 and β_2; i.e., this function is linear in the parameters β_1 and β_2, where $\cos(2\pi ft)$ and $\sin(2\pi ft)$ play the role of predictor variables.
- Adding an intercept term, say β_0, causes the following model

  $$Y_t = \beta_0 + \beta_1\cos(2\pi ft) + \beta_2\sin(2\pi ft) + X_t.$$

Remark for a linear trend

When we fit this model, we must be aware of the values used for the time t, as it has a direct impact on how we specify the frequency f. For example,

- if we have monthly data and use the generic time specification $t = 1, 2, \cdots, 12, 13, \cdots$, then we specify $f = 1/12$.
- if we have monthly data, but we use the years themselves as predictors; i.e., $t = 1990, 1991, 1992$, etc., we use $f = 1$, because 12 observations arrive each year.

Analyze example 3.6 by using a cosine trend model

- Parameter Estimates

Variable	DF	Estimate	SE	t-value	Pr > $\|t\|$
Intercept	1	14.80446	0.05624	263.24	< .0001
x_1	1	−2.23393	0.07953	−28.09	< .0001
x_2	1	−0.21791	0.07953	−2.74	.0070

- The fitted model (graphed in Figure 3.11) is

$$\hat{Y}_t = 14.80446 - 2.23393\cos(2\pi t) - 0.21791\sin(2\pi t).$$

- The residual process (depicted in Figure 3.12) is

$$\hat{X}_t = Y_t - \hat{Y}_t = Y_t - 14.80446 + 2.23393\cos(2\pi t) + 0.21791\sin(2\pi t).$$

- The residuals from the cosine trend fit appear to be somewhat stationary, but probably not white noise.

Figure 3.11 Beer sales data with a cosine trend model fit

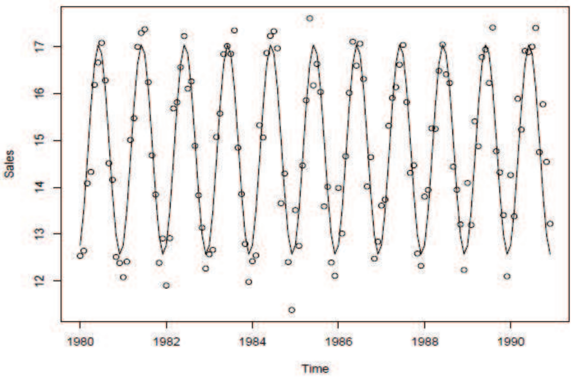


Figure 3.11: *Beer sales data with a cosine trend model fit.*

Figure 3.12. Residual from the cosine trend model fit

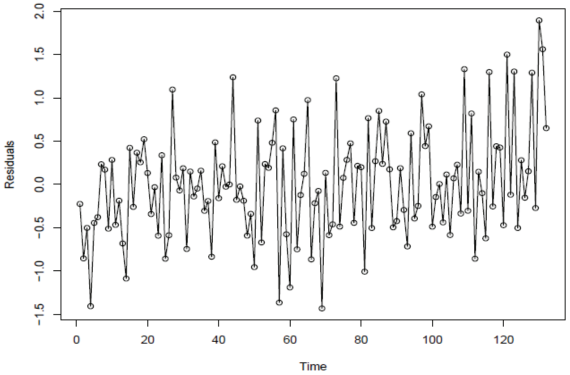


Figure 3.12: *Beer sales data. Residuals from the cosine trend model fit.*

Comparison between seasonal means and cosine trend model

The seasonal means and cosine trend models are competing models; that is, both models are useful for seasonal data.

- The cosine trend model is more parsimonious; i.e., it is a simpler model because there are 3 regression parameters to estimate. On the other hand, the (monthly) seasonal means model has 12 parameters that need to be estimated!
- This might not be a big deal if n is very large. It could be of consequence otherwise.
- A somewhat thorough mathematical argument on pp 36-39 (CC) should convince you of this result (if you are interested).

Summary for deterministic trends

- Consider the deterministic trend model

$$Y_t = \mu_t + X_t, \text{ with } \mathsf{E}(X_t) = 0.$$

- It is available to use the method of least squares to fit models with a trend of the following types:
  - a polynomial regression (Proc reg),
  - a trigonometric polynomial (Proc reg),
  - a seasonal means (Proc glm),
  - a linear combinations of these (Proc glm).
  - Why? These base on some large sample properties. Please read pp 40 (CC) to understand the reasons for above consideration or for further readings.

3.4. Interpreting regression output from SAS

In fitting the deterministic model

$$Y_t = \mu_t + X_t,$$

we have learnt the followings:

- if $\mathsf{E}(X_t) = 0$, for all t, least squares estimators are unbiased.
- if the variances of the least squares estimates (and standard errors) seen in SAS output is meaningful, we need $\mathsf{E}(X_t) = 0$, $\{X_t\}$ independent, and $\mathrm{var}(X_t) = \gamma_0$ (a constant). These assumptions are met if $\{X_t\}$ is a white noise process.
- If t tests and probability values are valid, we need the last three assumptions to be true and need the error process $\{X_t\}$ to be normal.

LSE of γ_0

- The matrix version of fitting the deterministic model is

$$\mathbf{Y} = \mu + \mathcal{E} = X\boldsymbol{\beta} + \mathcal{E}.$$

where $\mathbf{Y} = (Y_1, \cdots, Y_n)'$, $\mu = (\mu_1, \cdots, \mu_n)' = X\boldsymbol{\beta}$ and $\mathcal{E} = (X_1, \cdots, X_n)'$.

- If $\text{var}(X_t) = \gamma_0$ is constant, LSE of γ_0 is

$$\begin{aligned}\hat{\gamma}_0 &= \frac{1}{n-p}\mathbf{Y}'(I - P_X)\mathbf{Y} = \frac{1}{n-p}(\mathbf{Y} - \widehat{X\boldsymbol{\beta}})'(\mathbf{Y} - \widehat{X\boldsymbol{\beta}}) \\ &= \frac{1}{n-p}(\mathbf{Y} - \hat{\mu})'(\mathbf{Y} - \hat{\mu}) = \frac{1}{n-p}\sum_{t=1}^{n}(Y_t - \hat{\mu}_t)^2,\end{aligned}$$

where $\hat{\mu}_t$ is LSE of μ_t and $p = r(X)$.

Residual standard deviation: square root of LSE of γ_0

- The term $n-p$ is called the error degrees of freedom. If $\{X_t\}$ is independent, then $\hat{\gamma}_0$ is an unbiased estimator of γ_0.
- The residual standard deviation is defined by,

$$s = \sqrt{\hat{\gamma}_0} = \sqrt{\frac{1}{n-p}\sum_{t=1}^{n}(Y_t - \hat{\mu}_t)^2}.$$

  - The smaller s is, the better fit of the model. Therefore, in comparing two model fits (for two different models), we can look at the value of s in each model to judge which model may be preferred (caution is needed in doing this).
  - The larger s is, the noisier the error process is. This makes the least squares estimates more variable, and predictions less precise.

Analysis of variance or ANOVA table

- Decomposition of variation in a data set:

$$\mathbf{Y}'(I - P_1)\mathbf{Y} = \mathbf{Y}'(P_X - P_1)\mathbf{Y} + \mathbf{Y}'(I - P_X)\mathbf{Y}$$

or SST = SSR + SSE (sums of squares) or

$$\sum_{t=1}^{n}(Y_t - \overline{Y})^2 = \sum_{t=1}^{n}(\hat{Y}_t - \overline{Y})^2 + \sum_{t=1}^{n}(Y_t - \hat{Y}_t)^2.$$

- Sums of squares form the basis for (ANOVA) table

Source	df	SS	MS	F
Model	$p-1$	SSR	MSR$=\frac{\text{SSR}}{p-1}$	$F = \frac{\text{MSR}}{\text{MSE}}$
Error	$n-p$	SSE	MSE$=\frac{\text{SSE}}{n-p}$	
Total	$n-1$	SST		

Coefficient of determination R^2

- Coefficient of determination, proportion of the total variation in the data explained by the deterministic model, is

$$R^2 = \frac{\mathsf{SSR}}{\mathsf{SST}} = 1 - \frac{\mathsf{SSE}}{\mathsf{SST}}.$$

- The larger the R^2, the better the deterministic part of the model explains the variability in the data. Clearly,

$$0 \leq R^2 \leq 1.$$

Understanding coefficient of determination R^2

Critically, understand what R^2 does and does not measure. R^2 is computed under the assumption that the deterministic trend model is true and assesses how much of the variation in the data may be attributed to that relationship rather than just to inherent variation (maybe assumption is false).

- If R^2 is small, the random part contains a lot of variation in the data. Although the deterministic trend model is reasonable, it can only explain so much of the observed overall variation.
- A large R^2 (close to 1) does not imply that the model is the best model. In fact, R^2 could be very "high", but not relevant because a better model may exist.

Adjusted R^2

A slight variant of the coefficient of determination is the statistic

$$\overline{R}^2 = 1 - \frac{\mathsf{SSE}/(n-p)}{\mathsf{SST}/(n-1)}.$$

This is called the adjusted R^2 statistic. It is useful for comparing models with different numbers of parameters.

3.5. Residuals

- The fitted model is $\hat{Y}_t = \hat{\mu}_t$ and the residual process is

$$\hat{X}_t = Y_t - \hat{Y}_t.$$

  The residuals from the model fit are important. In essence, they serve as proxies (predictions) for the true errors X_t, which are not observed.

- The residuals can help us learn about the validity of the assumptions made in our model.

Standardized residuals

- If the model above is fit using least squares (and there is an intercept term in the model), then algebraically,

$$\sum_{t=1}^{n} \hat{X}_t = \sum_{t=1}^{n} (Y_t - \hat{Y}_t) = 0.$$

- Thus, the residuals have mean zero. Define unitless quantities, the standardized (studentized) residuals, by

$$\hat{X}_t^* = \hat{X}_t / s.$$

- If desired, use the standardized residuals for model diagnostic purposes.
- The standardized residuals defined here are not exactly zero mean, unit variance quantities, but they are approximately so.
- Thus, if the model is adequate, we would expect most standardized residuals to fall between -3 and 3.

3.5.1. Assessing normality

- If the error process $\{X_t\}$ is normally distributed, as a realization of $\{X_t\}$, the residuals should be approximately normally distributed.

- We can therefore diagnose this assumption by examining the (standardized) residuals and looking for evidence of normality.

- We can use histograms and normal probability plots (also known as quantile-quantile, or qq plots) and tests for normality to do this.

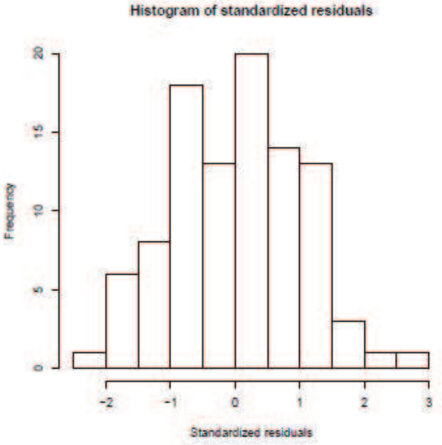
Histograms and qq plots to test normality

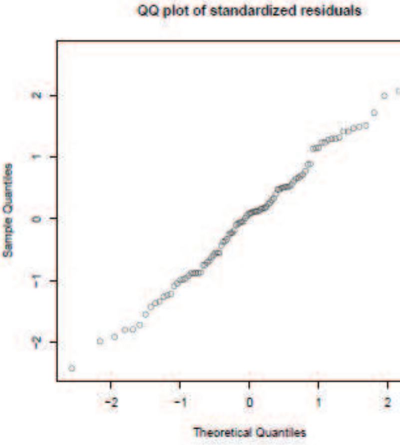
- Histograms which resemble heavily skewed empirical distributions are evidence against normality.
- A normal probability plot is a scatterplot of ordered residuals X_t (or standardized residuals $\hat{X}_t^*$) versus the ordered theoretical normal quantiles (or normal scores).
- The rationale behind this plot is simple. If the residuals are normally distributed, then plotting them versus the corresponding normal quantiles (i.e., values from a normal distribution) should produce a straight line (or at least close).

Histograms and qq plots for global temperature data

- We have fitted a straight line trend model to the global temperature data.
- Below are the histogram and qq plot for the standardized residuals. See sas program Normality3-4.sas
- Does normality seem to be supported?

Histograms and qq plots for global temperature data





Tests for normality

- Histograms and qq plots provide only visual evidence of normality.
- There are some tests for testing normality. They are
  - Shapiro-Wilk test (1965)
  - Kolmogorov-Smircling test
  - Cramér-von Mises test
  - Anderson-Darling test
  - Jarque-Bera test (1982) (in Autoreg procedure)

Details on Shapiro-Wilk test

- Shapiro-Wilk test statistic (1965) is defined as

$$W = \frac{(\sum_i^n a_i X_{(i)})^2}{\sum_i^n a_i (X_i - \bar{X})^2},$$

where

$$(a_1, \cdots, a_n) = \left(\frac{m' V^{-1}}{m' V^{-1} V^{-1} m} \right)^{1/2},$$

where $m = (m_1, \cdots, m_n)$ are the expected value of the order statistic of iid normal $N(0,1)$ and V is the covariance matrix of the order statistics.

Details on Shapiro-Wilk test

- These tests is a formal hypothesis test that can be used to test

$$H_0: \quad \text{the sresiduals are normally distributed}$$
$$H_1: \quad \text{the sresiduals are not normally distributed.}$$

- The higher this correlation, the higher the value of W. Small values of W are evidence against H_0.
- The distribution of W is complicated. But p-value can be calculated by Software.
- See SAS program Normality3-4.sas

3.5.2. Independence

- Plotting the residuals vs time can provide visual insight on whether or not the (standardized) residuals exhibit.
- Residuals that "hang together" are not what we would expect to see from a sequence of independent random variables.
- Similarly, residuals that oscillate back and forth too notably also do not resemble this sequence.

Standardized residuals from a linear trend fit for global temperature data

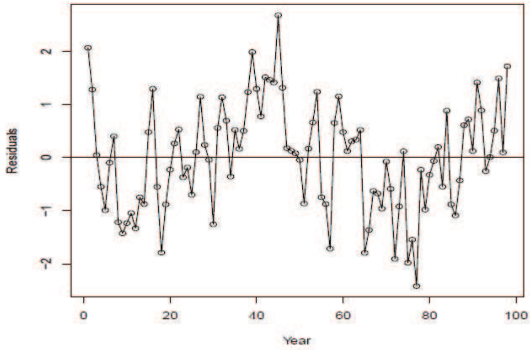


Figure 3.13: *Standardized residuals from the straight line trend model fit for the global temperature data. A horizontal line at zero has been added.*

Wald-Wolfowitz runs test for independence

- Wald-Wolfowitz runs test is a nonparametric test which calculates the number of runs in the (standardized) residuals.
- The formal test is

$$
\begin{aligned}
H_0: &\quad \text{the sresiduals are independent} \\
&\qquad\qquad\qquad \text{versus} \\
H_1: &\quad \text{the sresiduals are not independent.}
\end{aligned}
$$

Details on runs test for independence

In particular, the test examines the (standardized) residuals in sequence to look for patterns that would give evidence against independence. Runs above or below 0 (the approximate median of the residuals) are counted.

- A small number of runs would indicate that neighboring values are positively dependent and tend to hang together over time.
- Too many runs would indicate that the data oscillate back and forth across their median. This suggests that neighboring residuals are negatively dependent.
- Either too few or too many runs lead us to reject independence.

Distribution of runs

If the (standardized) residuals are truly independent, it is possible to write out the probability mass function $f_R(r)$ of R, the number of runs.

- If $r$is even, then

$$f_R(r) = C_{n_1-1}^{(r/2)-1} C_{n_2-1}^{(r/2)-1} / C_{n_1+n_2}^{n_1}$$

- if r s odd, then

$$f_R(r) = \left[C_{n_1-1}^{(r-1)/2} C_{n_2-1}^{(r-3)/2} + C_{n_1-1}^{(r-3)/2} C_{n_2-1}^{(r-1)/2} \right] / C_{n_1+n_2}^{n_1}.$$

- $n_1 =$ the number of sresiduals less than zero
- $n_2 =$ the number of sresiduals greater than zero
- $r_1 =$ the number of runs greater than zero
- $r_2 =$ the number of runs less than zero
- $r = r_1 + r_2$.

mean and variance of runs

- When n_1 and n_2 are large, the number of runs R is approximately normally distribution with mean

$$\mu_R = 1 + \frac{2n_1n_2}{n}$$

and variance

$$\sigma_R^2 = \frac{2n_1n_2(2n_1n_2 - n)}{n^2(n-1)}.$$

- Therefore, values of

$$Z = \frac{|R - \mu_R|}{\sigma_R} > z_{\alpha/2}$$

lead to the rejection of H_0.

Independence about global temperature data

A runs test on the standardized residuals produces

$$\text{observedruns} : 27, \text{expectedruns} : 49.81633, v\text{p-value} :< 0.0001.$$

- The p-value for the test is very small, so we would reject H_0.
- The evidence points to the sresiduals not being independent.
- The SAS output produces mean 49.81633 of runs under H_0, $n_1 = 52$ and $n_2 = 46$.

3.5.2. Sample autocorrelation function

- Consider the stationary stochastic process $\{Y_t\}$. The population ACF is

$$\rho_k = \text{corr}(Y_t, Y_{t-k}) = \gamma_k/\gamma_0,$$

where $\gamma_k = \text{Cov}(Y_t, Y_{t-k})$ and $\gamma_0 = \text{var}(Y_t)$.

- For a set of time series model $Y_1, Y_2, \cdots, Y_n$, the sample ACF, at lag k, is defined by

$$r_k = \frac{\sum_{t=k+1}^{n}(Y_t - \overline{Y})(Y_{t-k} - \overline{Y})}{\sum_{t=1}^{n}(Y_t - \overline{Y})^2},$$

where $\overline{Y}$ is the sample mean of $Y_1, Y_2, \cdots, Y_n$.

- The sample version r_k is a point estimator of ρ_k. For time series data $y_1, \cdots, y_n$, r_k is an estimate of ρ_k.

Sample ACF from standardized residuals

- The sample autocorrelation function of the (standardized) residual process $\{\hat{X}_t^*\}$ is defined by

$$r_k^* = \frac{\sum_{t=k+1}^{n} \left(\hat{X}_t^* - \overline{\hat{X}^*}\right)\left(\hat{X}_{t-k}^* - \overline{\hat{X}^*}\right)}{\sum_{t=1}^{n} \left(\hat{X}_t^* - \overline{\hat{X}^*}\right)^2}.$$

- When the sum of the standardized residuals equals zero (which occurs when least squares is used and when an intercept is included in the model), we have $\overline{\hat{X}^*} = 0$. The formula above reduces to

$$r_k^* = \frac{\sum_{t=k+1}^{n} \hat{X}_t^* \hat{X}_{t-k}^*}{\sum_{t=1}^{n} \left(\hat{X}_t^*\right)^2}.$$

Asymptotic properties of sample ACF from sresiduals

- Theorem: If the Sresidual process $\{\hat{X}_t^*\}$ is white noise, then

  (i) $\sqrt{n} r_k^* \xrightarrow{d} N(0,1)$. Or $r_k^* \sim \mathcal{AN}(0, 1/n)$ for large n.

  (ii) For $k \neq l$, $\text{Cov}(r_k^*, r_l^*) \approx 0$.

Comments on sample ACF from sresiduals

From the above theorem, we have the followings:

- If the Sresiduals are white noise, then

$$Pr(|r_k^*| \leq 2/\sqrt{n}) \geq 0.95,$$

  implying that r_k^* falls outside $[-2/\sqrt{n}, 2/\sqrt{n}]$ with a small probability. The $\pm 2/\sqrt{n}$ is the margin of error.

- If values of r_k^* are outside the margin of error, then they are not consistent with what we would see from a white noise process. This would suggest that there is dependence (autocorrelation) at lag k in the Sresidual process.

Correlogram

- The plot of r_k (or r_k^* if we are examining Sresiduals) versus k is called a correlogram.
- If we are assessing whether or not the process is white noise, it is helpful to put horizontal dashed lines at $\pm 2/\sqrt{n}$.
- We can easily see if the sample autocorrelations fall outside the margin of error $\pm 2/\sqrt{n}$.

Sample ACF from LTM fit for global temperature data

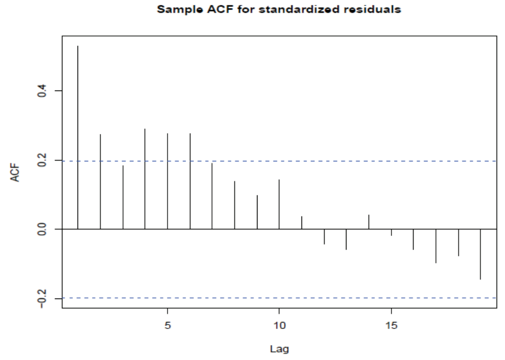


Figure 3.14: *Sample autocorrelation function for the standardized residuals from the straight line model fitted to the global temperature data.*

Analysis from sample ACF for global temperature data

- In Figure 3.14, we display the correlogram for the standardized sresiduals $\{\hat{X}_t^*\}$ from the straight line fit for the global temperature data.
- Many of the sample estimates r_k^* fall outside the $\pm 2/\sqrt{n}$ margin of error cutoff. It is clear that these residuals do not resemble a white noise process.
- There is actually quite a bit of structure left in the residuals. In particular, there is strong positive autocorrelation at early lags and the sample ACF tends to decay somewhat as k increases.

Exercise for sample ACF of two simulated N(0,1) series

- Let's take a moment and generate some white noise processes and examine their sample ACF.
- Figure 3.15 displays the results for two simulated white noise processes $e_t \sim \text{iid} N(0,1)$, where $n = 100$.
- With $n = 100$, the margin of error for the sample ACF r_k is $\pm 2/\sqrt{100} = \pm 0.2$.
- Figure 3.15 displays horizontal lines at the margin of error cutoffs.

Sample ACF for two simulated N(0,1) white noise series

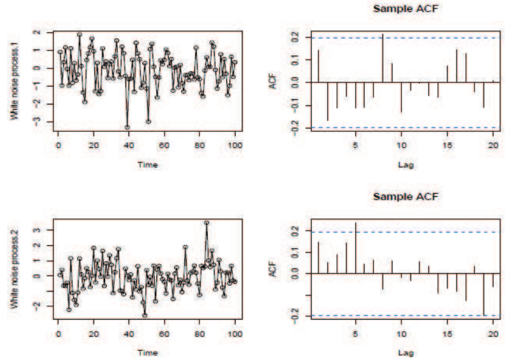


Figure 3.15: *Two simulated standard normal white noise processes with their associated sample autocorrelation functions.*

Analysis from sample ACF for two simulated N(0,1) series

Even though data are white noise, some r_k (for each realization) fall outside the margin of error $\pm 2/\sqrt{100} = \pm 0.2$.

- Every time we compare r_k to its margin of error cutoffs $\pm 2/\sqrt{n}$, we are testing $H_0 : \rho_k = 0$ at a significance level of approximately $\alpha = 0.05$.
- The upshot is that, on average 5 percent of time, we will observe a significant result which is really a "false alarm" (i.e., a Type I Error).
- When you are interpreting correlograms, you should be looking for definite patterns in r_k (especially at early lags).
- A stray statistically significant value of r_k at, say, lag $k = 14$ is probably just a false alarm.

Thank you for your attention!