Chapter 2. Time Series Regression and Exploratory Data Analysis

September 26, 2018

Outline

Basics

- Time series processes $\{x_t\}$ generally contain two different components:

  1. **systematic component** (that we would like to capture and model; e.g., trends, seasonality, etc.)

  2. **random component** (that is just inherent background noise in the process).

- Our goal as data analysts is to extract the systematic part of the data (and incorporate this into our model).

- If we do an adequate job of extracting the systematic part, then the only part left should be random variation, which can hopefully be modeled by white noise.

Deterministic vs Stochastic Trends

- The simulated random walk in Exhibit 2.1 shows an "upward trend", which is caused by a strong correlation between the nearby data points. The true process is $x_t = x_{t-1} + w_t = \sum_{i=1}^{t} w_t$. We call such a trend a **stochastic (random) trend**.
- The average monthly temperature series in Exhibit 1.7 shows a cyclical trend. We may model it by $x_t = \mu_t + y_t$, where μ_t is a deterministic function with period 12 ($\mu_t = \mu_{t-12}$), and y_t is a zero mean unobserved variation (or noise). We say that this model has a **deterministic (fixed) trend**.
- Stochastic trend can be mistakenly identified as a deterministic when it is not like in Exhibit 2.1.
- Deterministic trend can be hard to detect due to much background noise.

Deterministic Trend Models

- In this chapter, we consider models of the form

$$x_t = \mu_t + y_t$$

  where μ_t is a *deterministic* function that describes the trend and y_t is a *random error*.

- If $E(y_t) = 0$ for all t then

$$E(x_t) = \mu_t$$

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

Preprocessing nonstationarity in the mean

1. Detrending: Estimate μ_t with $\widehat{\mu}_t$ and then subtract it from the data. Then model the residuals as

$$\widehat{y}_t = x_t - \widehat{\mu}_t$$

as a stationary process.

- We can use *regression methods* to estimate μ_t and perform standard diagnostics on the residuals $\widehat{y}_t$.
- One popular choice is

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

which is a kth order polynomial in time.

- If deterministic trend is cyclical then we might consider

$$\mu_t = \beta_0 + \sum_{j=1}^{m} [\alpha_j \cos(\omega_j t) + \beta_j \sin(\omega_j t)].$$

- We then forecast by using the residual process $\widehat{y}_t$ and inverting the procedures above to arrive at forecasts for $\{x_t\}$.

Constant Trend Model

- We first consider a model with a constant mean function (e.g. $\{y_t\}$ is stationary):

$$x_t = \mu + y_t, \qquad E(y_t) = 0.$$

- We may estimate μ by

$$\overline{x} = \frac{1}{n}\sum_{t=1}^{n} x_t. \qquad \text{(unbiased)}$$

- If $\{x_t\}$ is *stationary* with autocorrelation function ρ_h, then Ex 2.17 (CaC or pp27 on SaS) shows that

$$Var\overline{x} \;=\; \frac{\gamma_0}{n}\left\{1 + 2\sum_{h=1}^{n-1}\left(1 - \frac{h}{n}\right)\rho_x(h)\right\}.$$

Constant Trend Model

- If $\{x_t^*\}$ is an iid process then $Var\ \overline{x}^* = \frac{\gamma_0}{n}$.
- If $\rho_1 \neq 0$ but $\rho_h = 0$ for $h > 1$, then $Var\ \overline{x} \sim \frac{\gamma_0}{n}\{1 + 2\rho_1\}$ for n-large.
- Depending on ρ_h above, $\overline{x}$ can be better or worse than $\overline{x^*}$.

Moving Average (1)

- Let $x_t = w_t + \theta w_{t-1}$, then $\rho_1 = \frac{\theta}{1+\theta^2}$ and $\rho_h = 0$ for $h > 1$.

- We have

$$Var\ \overline{x} = \frac{\gamma_0}{n}\left[1 + \frac{2\theta}{1+\theta^2}\left(\frac{n-1}{n}\right)\right] \approx \frac{(1+\theta)^2}{1+\theta^2}\frac{\gamma_0}{n}.$$

- For $-1 < \theta < 0$, say $\theta = -0.5$, $Var\ \ \overline{x} \approx \frac{0.2\gamma_0}{n} < \frac{\gamma_0}{n}$. The negative correlation ρ_1 improves the estimator.

- For $0 < \theta < 1$, say $\theta = +0.5$, $Var\ \overline{x} \approx \frac{1.8\gamma_0}{n} > \frac{\gamma_0}{n}$. The positive correlation ρ_1 makes the estimator less precise.

Autoregressive (1) or AR(1):

- For a first order autoregressive process

$$x_t = \phi x_{t-1} + w_t, \qquad -1 < \phi < 1,$$

- We can show that $\rho_h = \phi^{|h|}$ for all h.
- Then

$$Var\ \overline{x} \approx \frac{\gamma_0}{n} \sum_{h=-\infty}^{\infty} \phi^{|h|} = \frac{(1+\phi)}{(1-\phi)} \frac{\gamma_0}{n}.$$

- This shows that when $\phi \to 1^-$, the variance of $\overline{x}$ dramatically increases. In fact, $\phi = 1$ corresponds to the simple random walk process.
- The variance of sample mean can be very large or different for a nonstationary process.

Classical Regression

- **Why regression?** It is useful for forecasting and smoothing. It makes it easy to include external data.

- Let $z_t = (1, z_{t1}, \ldots, z_{tq})^{'}$ be a collection of possible inputs (fixed/constant covariates) and $\beta = (\beta_0, \beta_1, \ldots, \beta_q)^{'}$ be the vector of coefficients

- For a time series x_t,

$$x_t = \beta^{'} z_t + w_t, \qquad w_t \textit{ iid } N(0, \sigma^2).$$

- Use `lm(x~z, data=dataname)` in R.

Classical Regression

- **Monthly Chicken Prices:** From Aug 2001 to July 2016. We fit a model

$$x_t = \beta_0 + \beta_1 z_t + w_t, \qquad z_t = 2001\frac{7}{12}, 2001\frac{8}{12}, \ldots, 2016\frac{6}{12}$$

  where w_t *iid* $N(0, \sigma_w^2)$.

- The trend estimate is

$$\widehat{\mu}_t = -7131 + 3.59 z_t.$$

- There is a significant estimated increase of \$3.59 cents per year.
- Does it have a good fit?

Classical Regression

- You compare models using R^2, AIC, and BIC among others.
- **Akaike Information Criterion**

$$AIC = \log\left(\hat{\sigma}_k^2\right) + \frac{n + 2k}{n}$$

where k is the number of parameters in the model.

- We penalize the error variance with a term proportional to k.
- **Bias corrected AIC (AICc)**

$$AICc = \log\left(\hat{\sigma}_k^2\right) + \frac{n + k}{n - k - 1}$$

where k is the number of parameters in the model.

- AIC with a correction for finite sample sizes.
- **Bayesian/Schwarz Information Criterion**

$$BIC = \log\left(\hat{\sigma}_k^2\right) + \frac{k \log n}{n}.$$

Classical Regression

- **Pollution(P), Temperature (T), and Mortality(M) data (n=508):** The following models are fitted.

$$
\begin{aligned}
M_t &= \beta_0 + \beta_1 t + w_t \qquad (2.18)\\
M_t &= \beta_0 + \beta_1 t + \beta_2(T_t - \overline{T}) + w_t \qquad (2.19)\\
M_t &= \beta_0 + \beta_1 t + \beta_2(T_t - \overline{T}) + \beta_3(T_t - \overline{T})^2 + w_t \qquad (2.20)\\
M_t &= \beta_0 + \beta_1 t + \beta_2(T_t - \overline{T}) + \beta_3(T_t - \overline{T})^2 + \beta_4 P_t + w_t \quad (2.21)
\end{aligned}
$$

Table 2.2. Summary Statistics for Mortality Models

Model	k	SSE	df	MSE	R^2	AIC	BIC
(2.18)	2	40,020	506	79.0	.21	5.38	5.40
(2.19)	3	31,413	505	62.2	.38	5.14	5.17
(2.20)	4	27,985	504	55.5	.45	5.03	5.07
(2.21)	5	20,508	503	40.8	.60	4.72	4.77

Lab for Today

- Fit a cubic polynomial (centered) model to the chicken data and see if it improves the fit. Have the data and the cubic fit in one plot.

- Generate a signal $x_t = 1 + 3t + e_t$, with $n = 200$ and where
  1) $e_t \sim N(0, 100)$;
  2) $e_t \sim 0.3w_t - 0.3w_{t-1} + 0.4w_{t-2}$.

- For 1 and 2 above, estimate and remove the trend. Examine the acf of the residuals.

Preprocessing nonstationarity in the mean

2. Differencing: Developed by Box and Jenkins, is to apply differencing repeatedly to the series $\{x_t\}$ until the differenced observations resemble a realization of a stationary time series.

- We can then use the theory of stationary processes for the modeling, analysis, and prediction of the differenced data.

- Transform back in terms of the original series $\{x_t\}$.

Differencing

- **Random Walk Process:** $x_t = \sum_{j=1}^{t} w_j, E(x_t) = 0$ but $var(x_t) = s\sigma_w^2 = \gamma_{t,s}, s \leq t$.
- Consider the first-order difference process, i.e., $\nabla x_t = x_t - x_{t-1} = w_t$.
- The **first difference process** $\{\nabla x_t\}$ is stationary.

Backshift Operator

- *Backshift operator* B is defined as

$$Bx_t = x_{t-1}.$$

- Higher orders: $B^2 x_t = x_{t-2}, \ldots, B^k x_t = x_{t-k}$
- The first difference is $\nabla x_t = x_t - x_{t-1}$; clearly,

$$\nabla x_t = (1 - B) x_t$$

- A *difference of order d* is defined as

$$\nabla^d = (1 - B)^d.$$

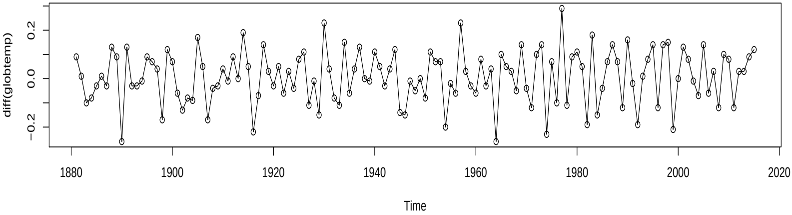
  A *first-order* $(d = 1)$ difference can eliminate a linear trend.
- For $d = 2$,

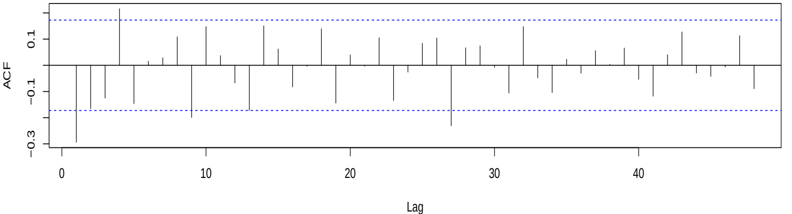
$$\nabla^d = (1 - B)^d x_t = (1 - 2B + B^2) x_t = x_t - 2x_{t-1} + x_{t-2}.$$

  which can eliminate a quadratic trend, and so on.

Data Exploration

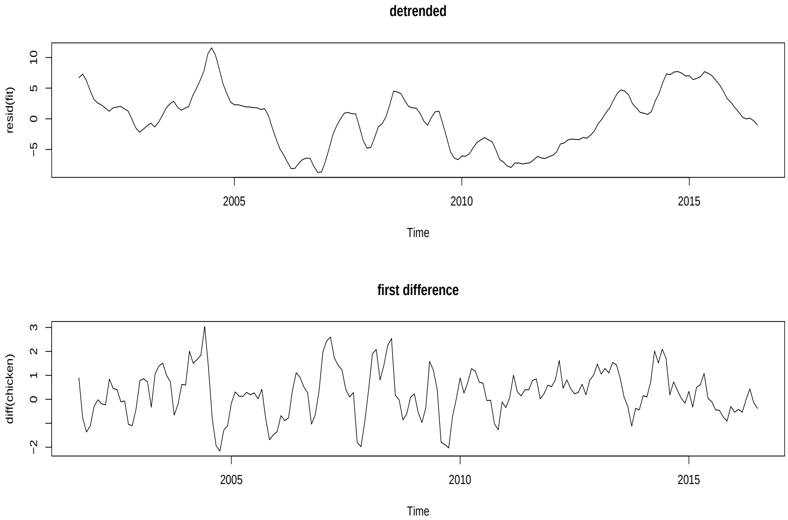
- **Global Temp.** The differenced global temperature data and its sample ACF.





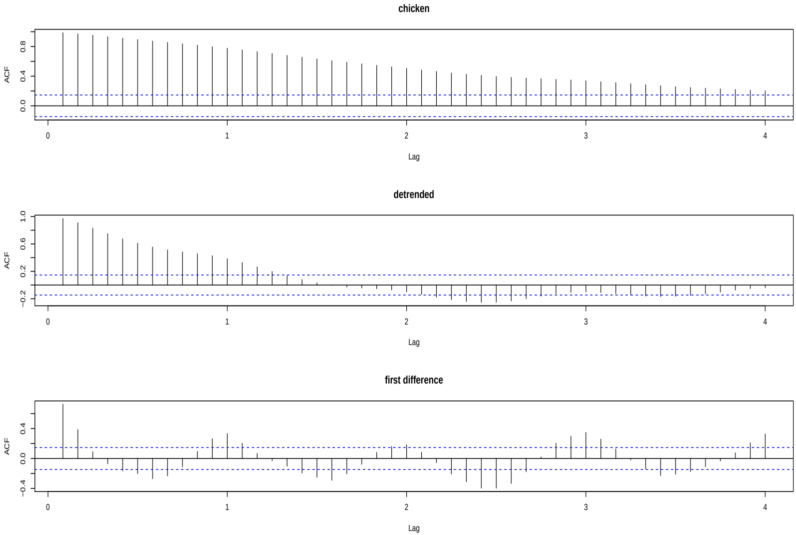
Data Exploration

- **Chicken Prices:** The detrended model is $\hat{y}_t = x_t + 7131 - 3.59t$.
- Detrended and differenced (first oder) chicken price data



Data Exploration

- **Chicken Prices cont'd.** Sample ACFs of the original data, the detrended and the differenced data.



Data Exploration

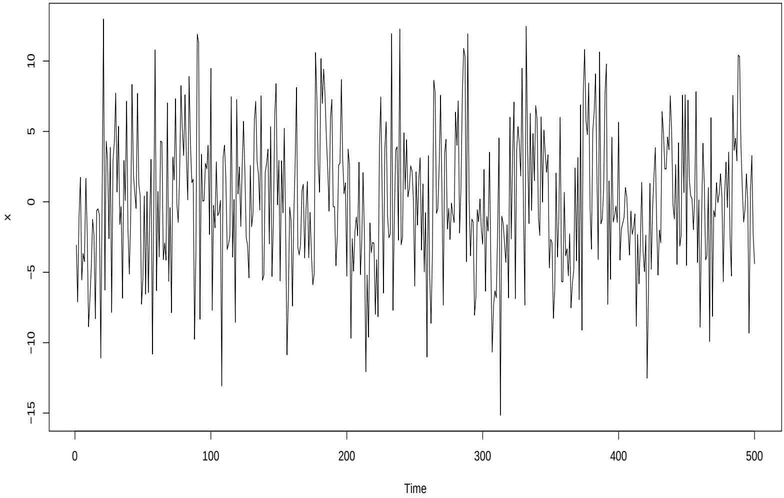
- The most commonly used is the classical Box-Cox transformation family of the form

$$y_t = \begin{cases} (x_t^\lambda - 1)/\lambda & \lambda \neq 0 \\ \log x_t & \lambda = 0. \end{cases}$$

- **E.g.** Glacial varve (yearly thickness sedimentary deposits) series from one location in Massachusetts for 634 years exhibits a non-constant variance that seems to change with time. See R.
- **E.g.** Stock data: Dow Jones, S&P 500, etc.
- **E.g.** The `oil` and `gas` data.

Data Exploration

- **Discovering a signal from a noise:**



Smoothing

- **Moving Average Smoother:** If we want to *filter* out long-term trend and seasonal components of x_t then we can use the symmetric moving average

$$m_t = \sum_{j=-k}^{k} a_j x_{t-j},$$

where $a_j = a_{-j} \geq 0$ and $\sum_{j=-k}^{k} a_j = 1$.

- Use `filter` in R.

- **SOI data:** Using $a_0 = a_{\pm 1} = \cdots = a_{\pm 5} = 1/12$ and $a_{\pm 6} = 1/24$; $k = 6$, the El Ni$\widetilde{n}$o effect is highlighted although it is still a bit rough.

Smoothing

- **Kernel Smoothing (Nadaraya-Watson):** Uses a kernel/weight function and is given as

$$m_t = \sum_{i=1}^{n} w_i(t) x_j,$$

where

$$w_i(t) = K\left(\frac{t-i}{b}\right) \Big/ \sum_{j=1}^{n} K\left(\frac{t-j}{b}\right)$$

are the weights and $K(\cdot)$ is a kernel function.
- Use `ksmooth` in R and play with the bandwidth b.

- **SOI data:** Using a normal kernel

$$K(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-z^2/2\right),$$

and depending on the bandwidth parameter, we can *smooth* out some more (or not) the El Niño effect.

Smoothing

- **Smoothing Splines or Roughness Penalty:** It calculates a cubic spline m_t as the minimizer of

$$\sum_{i=1}^{m} [x_t - m_t]^2 + \lambda \int (m_t^{''})^2 dt = \text{ fit } + \text{roughness penalty}$$

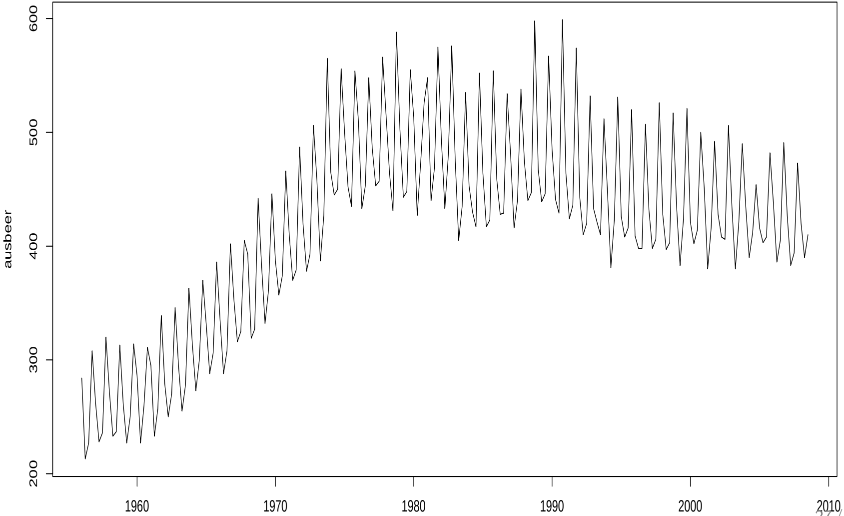
  where λ is the smoothing parameter that balances between *data fit* and *smoothness*.

- Use `smooth.spline` in R and play with λ.

- **(LOcally WEighted Scatterplot Smoothing (lowess):** Local polynomial/nearest neighbor regression.Uses nearest k neighbors to predict x_t via regression. Use `lowess` in R.

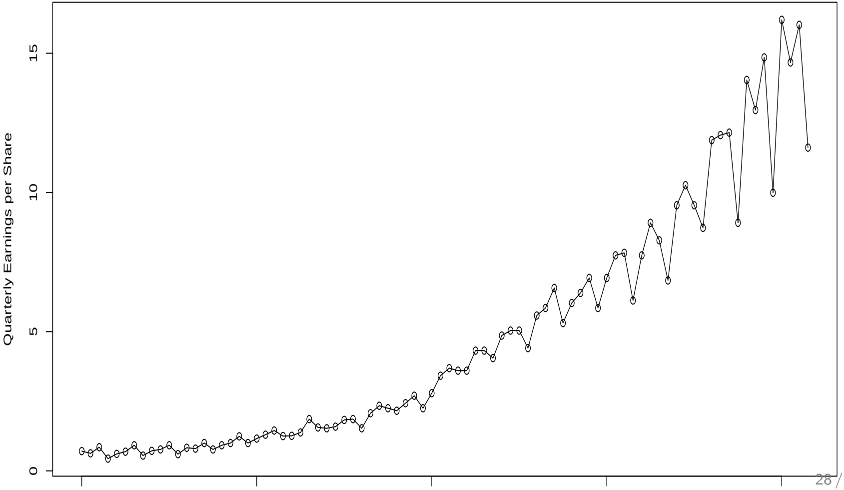
Signal Decomposition in R

- **Structure 1:** $x_t = \mathit{Trend} + \mathit{Seasonal} + \mathit{Random}$ **(**Additive)
- Quarterly beer production in Australia (in megalitres) from 1956:Q1 to 2008:Q3



Signal Decomposition in R

- **Structure 2:** $x_t = \mathit{Trend} \times \mathit{Seasonal} \times \mathit{Random}$ **(Multiplicative)**
- Johnson and Johnson quarterly earnings per share, 84 quarters (21 years) measured from the first quarter of 1960 to the last quarter of 1980.



Signal Decomposition in R

Trend estimator for `decompose`

- For **quarterly** data:

$$\frac{1}{8}x_{t-2} + \frac{1}{4}x_{t-1} + \frac{1}{4}x_t + \frac{1}{4}x_{t+1} + \frac{1}{8}x_{t+2}.$$

- For **monthly** data:

$$\frac{1}{24}x_{t-6} + \left(\sum_{j=-5}^{5} \frac{1}{12}x_{t+j}\right) + \frac{1}{24}x_{t+6}$$

Signal Decomposition in R

Seasonal effect estimator for `decompose`

- Seasonal effects are estimated using the de-trended series. *E.g.*, estimating a quarterly effect for quarterly data. Or estimating a monthly effect for monthly data.

- Average the de-trended values for a specific season. *E.g.*, To get a seasonal effect for a particular quarter/month, we average the de-trended values for that particular quarter/month in the entire series.

- Random = Series - Trend - Season for additive model , and Random = Series / (Trend$\times$Season) for the multiplicative model.

Signal Decomposition in R

De-seasonalizing or seasonal adjustment

- Seasonal effects are used to adjust future values.
- **E.g.** Suppose the 2008 4th quarter production for Australian beer is 435. The seasonally adjusted or de-seasonalized value is 435 - 69.269=365.73.
- **E.g.** Suppose the 2008 3rd quarter earning for J&J is 13. The seasonally adjusted or de-seasonalized value is 13/ 1.114= 11.67.

Summary

- Systematic vs random components
- Data exploration and transformations (detrending, differencing, Box-Cox)
- Classical and non-parametric regression (smoothing)
- `decomposition` in R

Lab for Today

- Generate a simple random walk data. Apply first-order differencing. Did it remove non-stationarity? Apply second-order differencing. Did it remove non-stationarity?

- Consider the immigration data from BB. Stationary? If not stationary then try to stationarize the data.

- Consider the Monthly Australian Beer Consumption data from BB. Decompose the data and interpret.