Lecture 25: Correlated Errors

Pratheepa Jeganathan

11/18/2019

Recap

- What is a regression model?
- Descriptive statistics – graphical
- Descriptive statistics – numerical
- Inference about a population mean
- Difference between two population means
- Some tips on R
- Simple linear regression (covariance, correlation, estimation, geometry of least squares)
  - Inference on simple linear regression model
  - Goodness of fit of regression: analysis of variance.
  - F-statistics.
  - Residuals.
  - Diagnostic plots for simple linear regression (graphical methods).

Recap

- Multiple linear regression
  - Specifying the model.
  - Fitting the model: least squares.
  - Interpretation of the coefficients.
  - Matrix formulation of multiple linear regression
  - Inference for multiple linear regression
    - T-statistics revisited.
    - More F statistics.
    - Tests involving more than one β.
- Diagnostics – more on graphical methods and numerical methods
  - Different types of residuals
  - Influence
  - Outlier detection
  - Multiple comparison (Bonferroni correction)
  - Residual plots:
    - partial regression (added variable) plot,
    - partial residual (residual plus component) plot.

Recap

- Adding qualitative predictors
  - Qualitative variables as predictors to the regression model.
  - Adding interactions to the linear regression model.
  - Testing for equality of regression relationship in various subsets of a population
- ANOVA
  - All qualitative predictors.
  - One-way layout
  - Two-way layout
- Transformation
  - Achieving linearity
  - Stabilize variance
  - Weighted least squares

Correlated Errors

Outline

- Today, we will consider another departure from our usual model for the error variance (i.e. equal variance σ^2 and independent).
- Before we do this, let's recall *weighted least squares* method.

Weighted least squares (WLS)

- In the last set of notes, we considered a model

$$Y = X\beta + \epsilon, \qquad \epsilon \sim N(0, W^{-1})$$

where

$$W^{-1} = \sigma^2 \begin{pmatrix} V_1 & 0 & 0 & \cdots & 0 \\ 0 & V_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & V_n \end{pmatrix},$$

and

$$\sigma^2 V_i = \text{Variance}\left(\epsilon_i\right).$$

- This model has independent errors, but of different variance: a *heteroscedastic* model.

The fix

- We saw that by defining

$$\tilde{Y} = W^{1/2}Y, \qquad \tilde{X} = W^{1/2}X$$

  we transformed our original model to more familiar model:

$$\tilde{Y} = \tilde{X}\beta + \varepsilon, \qquad \varepsilon \sim N(0, \sigma^2 I).$$

- The usual estimator in this model is the *WLS* estimator

$$\hat{\beta}_{\mathrm{WLS}} = (X^T W X)^{-1} X^T W Y.$$

The implications

- If we ignore *heteroscedasticity* then our *OLS* estimator has distribution

$$\hat{\beta} = (X^T X)^{-1} X^T Y \sim \mathrm{N}(\beta, \sigma^2 (X^T X)^{-1} X^T W^{-1} X (X^T X)^{-1}).$$

- This form of the variance matrix is called the *sandwich form*.
- **This means that our `Std. Error` column will be off! In other words, `R` will report t statistics that are off by some multiplicative factor!**
- Another reason to worry about W is that if we use the correct W, we have a more *efficient* unbiased estimator: smaller confidence intervals.

The implications

Using the correct W *proportional to inverse variance of the errors* and form the WLS estimator we have

$$\hat{\beta}_{\mathsf{WLS}} \sim \mathsf{N}(\beta, \sigma^2(X^T W X)^{-1}).$$

Autocorrelation

- The model of the variance that we will consider today is a model where the errors are *correlated*.
- Common examples of this type of errors occur in time series data, a common model for financial applications.
- Why should we worry?
    - Just as in the *heteroscedastic case*, ignoring autocorrelation can lead to underestimates of `Std. Error` $\rightarrow$ inflated t's $\rightarrow$ false positives.

What is autocorrelation?

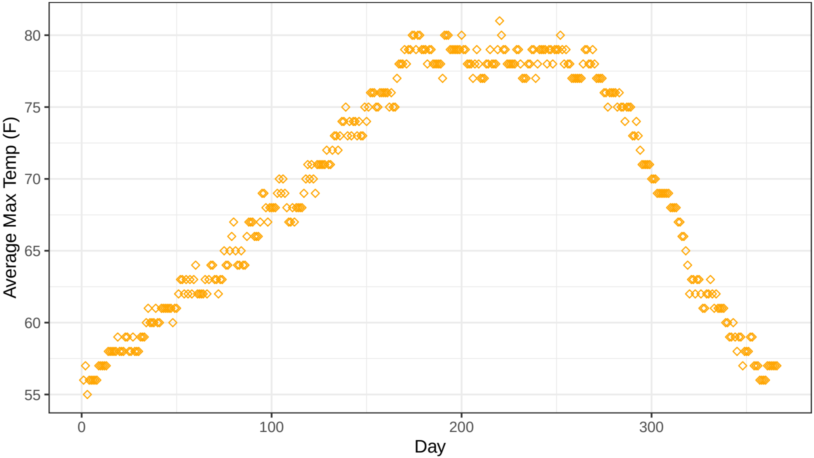
- Suppose we plot Palo Alto's daily average temperature – clearly we would see a pattern in the data.
- Sometimes, this pattern can be attributed to a deterministic phenomenon (i.e. predictable seasonal fluctuations).
- Other times, "patterns" are due to correlations in the noise, maybe small time fluctuations in the stock market, economy, etc.
  - Example: financial time series: NASDAQ close prices.
  - Example: residuals regressing consumer expenditure on money stock (this one is discussed in your textbook and used as an example below).
- Sometimes, this pattern can attribute to omission of a variable that should be in the model.

Average Maximum Temperature in Palo Alto

- The daily max temperature shows clear seasonal fluctuations.

```
PA.temp = read.table('http://stats191.stanford.edu/data/pal
  header=F, skip=2)
p = ggplot(PA.temp) +
  geom_point(aes(x = seq(1, nrow(PA.temp)),y = V3),
    color = "orange", shape = 23) +
  theme_bw() +
  ylab("Average Max Temp (F)") +
  xlab("Day")
```

Average Maximum Temperature in Palo Alto



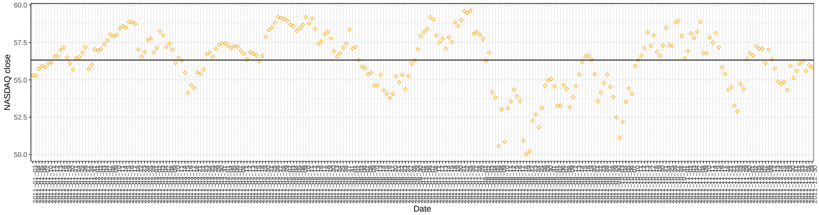
NASDAQ daily close 2011

- Another example of a time series can be found from financial data. The price of many assets fluctuate from day to day.
- Still, there is a *pattern* in this process.
- Given enough information, we might try to also explain this pattern as a deterministic model, like the temperature data. (This is, in some sense, what business news sites try to do on a daily basis).
- A simpler model for this pattern is that of some unexplainable noise. . .
- Below, we plot some closing prices of NASDAQ for the year 2011. Data was obtained from on yahoo finance.

NASDAQ daily close 2011

```
fname = 'http://stats191.stanford.edu/data/nasdaq_2011.csv
nasdaq.data = read.table(fname,
  header=TRUE, sep=',')
nasdaq.p = ggplot(nasdaq.data) +
  geom_point(aes(x = Date,y = Close), color = "orange",
    shape = 23) +
  theme_bw() +
  ylab("NASDAQ close") +
  xlab("Date") +
  geom_hline(yintercept = mean(nasdaq.data$Close)) +
  theme(axis.text.x = element_text(angle = 90))
```

NASDAQ daily close 2011



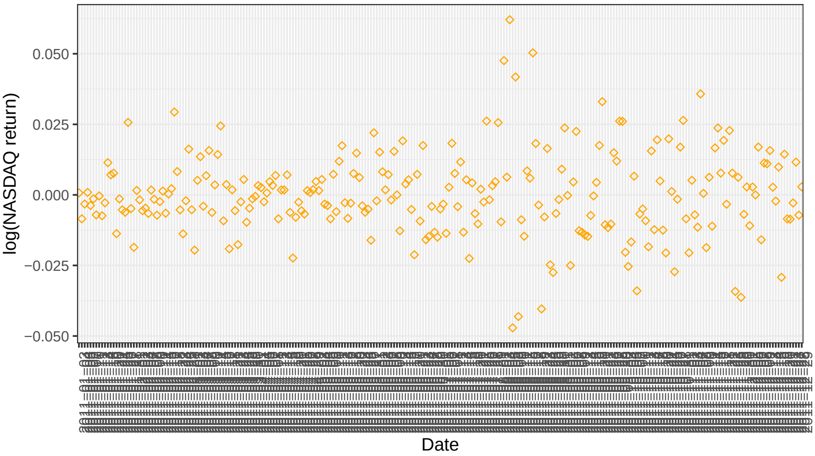
NASDAQ daily close 2011

- Let's look at the plot of log(NASDAQ return)
- Let X_t be NASDAQ close
- log(NASDAQ return) $= \log\left(X_t / X_{t-1}\right)$

```
ndays = length(nasdaq.data$Date)
log_return = log(nasdaq.data$Close[2:ndays] /
    nasdaq.data$Close[1:(ndays-1)])
df = data.frame(Date =
    nasdaq.data$Date[2:nrow(nasdaq.data)],
  log_return = log_return)
log.nasdaq.return.p = ggplot(df) +
  geom_point(aes(x = Date,y = log_return),
    color = "orange",
    shape = 23) +
  theme_bw() +
  ylab("log(NASDAQ return)") +
  xlab("Date") +
  theme(axis.text.x = element_text(angle = 90))
```

NASDAQ daily close 2011

```
log.nasdaq.return.p
```



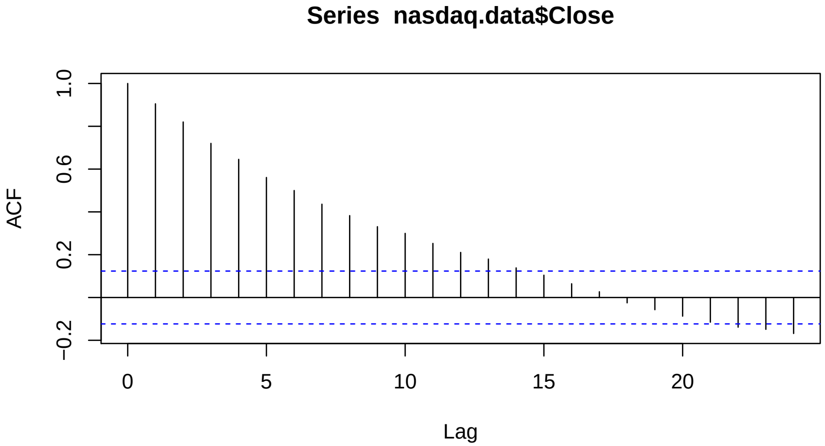
NASDAQ daily close 2011, ACF

- One way this noise is measured is through the *ACF (Auto-Correlation Function)*, which we will define below.
- A time series with no auto-correlation (i.e. our usual multiple linear regression model) has an ACF that contains only a spike at 0.

NASDAQ daily close 2011, ACF

- The NASDAQ close clearly has some auto-correlation.

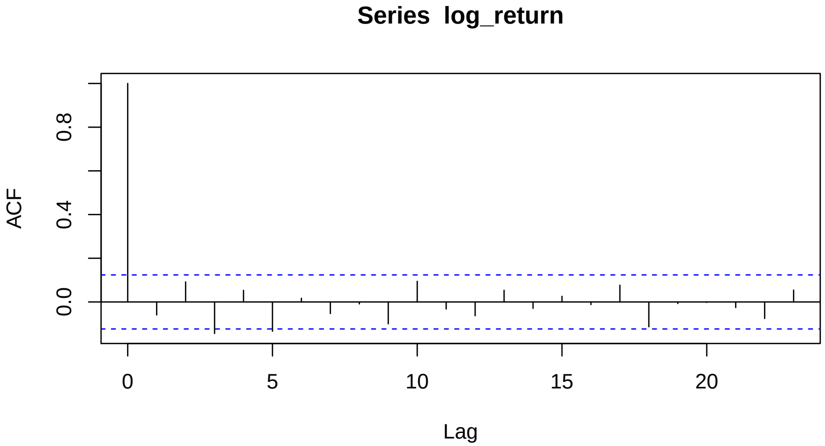
```r
acf(nasdaq.data$Close)
```



NASDAQ daily log return 2011, ACF

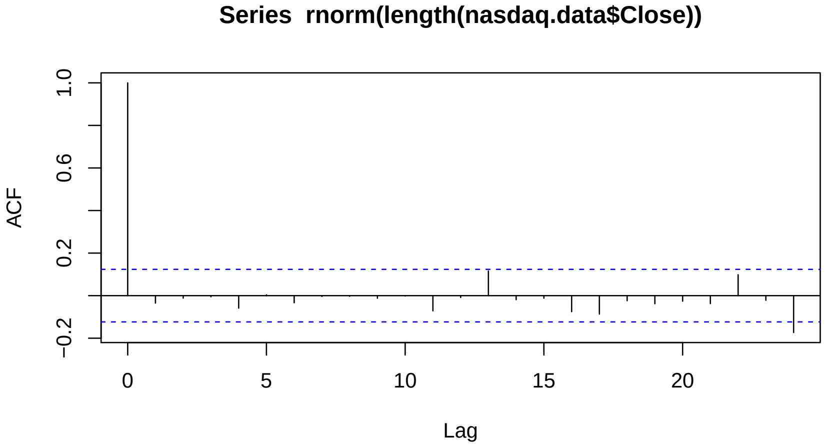
- The log NASDAQ return shows no auto-correlation.

```r
acf(log_return)
```



ACF of independent noise

```r
acf(rnorm(length(nasdaq.data$Close)))
```



Expenditure vs. stock (**CH** Chapter 8.2, Page 210)

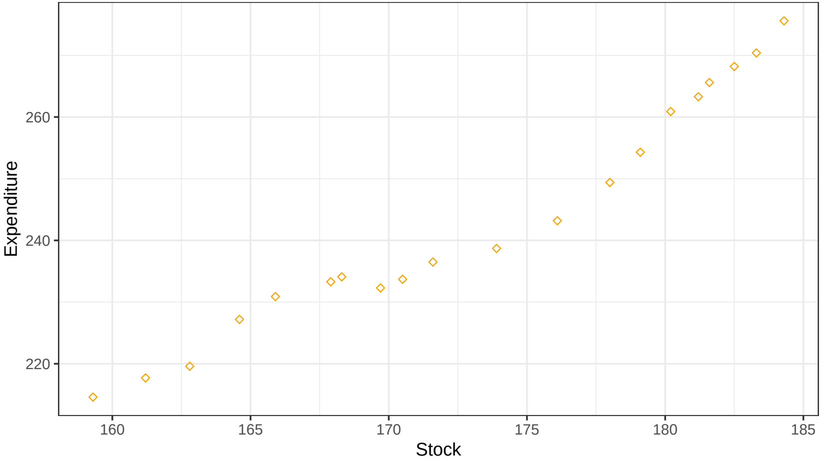
- The example we will consider is that of *consumer expenditure* vs. *money stock*, the supply of available money in the economy.
- Data are collected yearly, so perhaps there is autocorrelation in the model

$$\texttt{Expenditure}_t = \beta_0 + \beta_1 \texttt{Stock}_t + \epsilon_t$$

Expenditure vs. stock

```
fname = 'http://stats191.stanford.edu/data/expenditure.tabl
expenditure.table =read.table(fname,
  header=T)
expenditure.p = ggplot(expenditure.table) +
  geom_point(aes(x = Stock,y = Expenditure),
    color = "orange",
    shape = 23) +
  theme_bw() +
  ylab("Expenditure") +
  xlab("Stock")
```

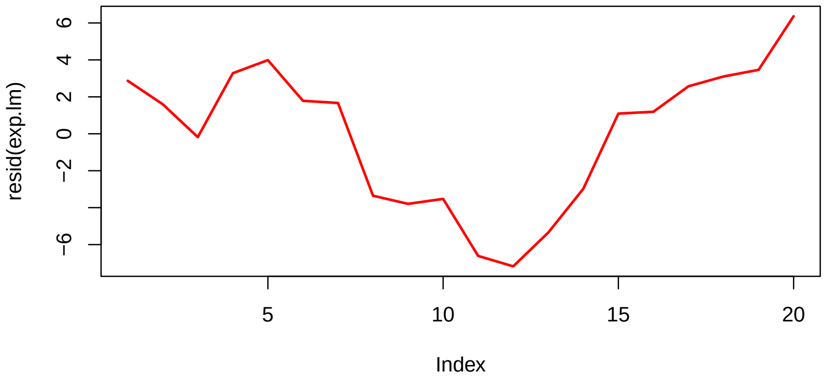
Expenditure vs. stock



Expenditure vs. stock: residuals

- A plot of residuals against `time`, i.e. their index may show evidence of autocorrelation.

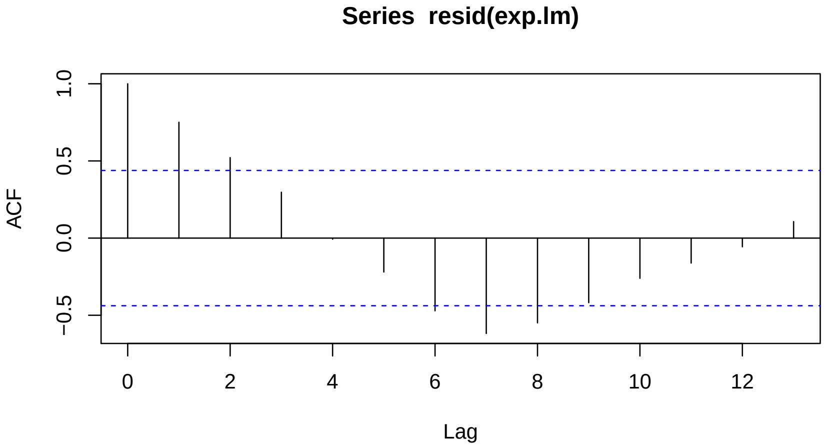
```
exp.lm = lm(Expenditure ~ Stock,
  data = expenditure.table)
plot(resid(exp.lm), type='l', lwd=2, col='red')
```



ACF of residuals

- A plot of the ACF may also help. Since there seem to be some points outside the confidence bands, this is some evidence that auto-correlation is present in the errors.

```
acf(resid(exp.lm))
```



Models for autocorrelated errors

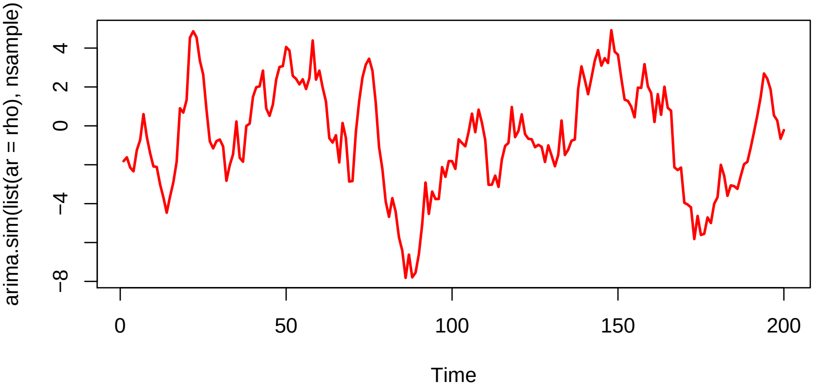
- AR(1) noise (Autorrgressive with order 1) noise
  - Suppose that, instead of being independent, the errors in our model were

    $$\varepsilon_t = \rho \cdot \varepsilon_{t-1} + \omega_t, \qquad -1 < \rho < 1$$

    with $\omega_t \sim N(0, \sigma^2)$ independent.
  - If ρ is close to 1, then errors are very correlated, $\rho = 0$ is independence.
  - This is "Autoregressive Order (1)" noise [AR(1)].
- Many other models of autocorrelation exist: ARMA (autoregressive moving average), ARIMA (autoregressive integrated moving average), ARCH (Autoregressive Conditionally Heteroskedastic), GARCH (Generalized Autoregressive Conditionally Heteroskedastic), etc.

AR(1) noise, $\rho = 0.9$

```r
nsample = 200
rho = 0.95
mu = 1.0
plot(arima.sim(list(ar=rho), nsample),
  lwd=2, col='red')
```



Autocorrelation function

- For a "stationary" time series $(Z_t)_{1 \leq t \leq \infty}$ define

$$ACF(t) = \text{ Cor}(Z_s, Z_{s+t}).$$

- Stationary means that correlation above does not depend on s.
- For AR(1) model,

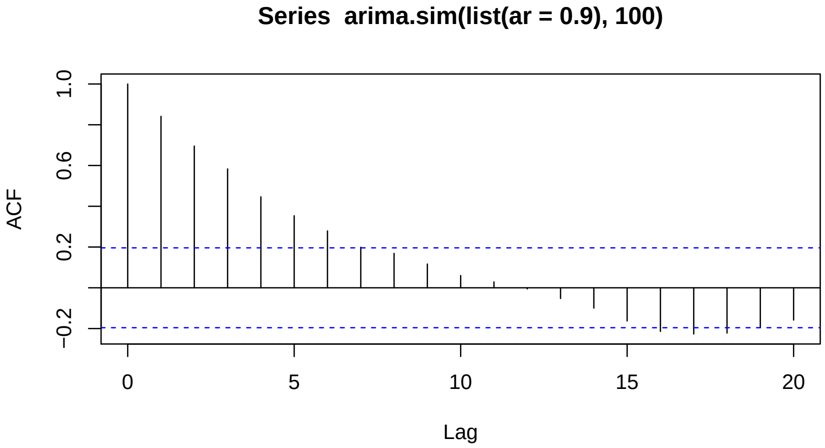
$$ACF(t) = \rho^t.$$

- For a sample $(Z_1, \dots, Z_n)$ from a stationary time series

$$\widehat{ACF}(t) = \frac{\sum_{j=1}^{n-t}(Z_j - \overline{Z})(Z_{t+j} - \overline{Z})}{\sum_{j=1}^{n}(Z_j - \overline{Z})^2}.$$

ACF of AR(1) noise, $\rho = 0.9$

```
acf(arima.sim(list(ar=0.9), 100))
```



Effects on inference

- So far, we have just mentioned that things *may* be correlated, but not thought about how it affects inference.
- Suppose we are in the "one sample problem" setting and we observe

  $$W_i = Z_i + \mu, \qquad 1 \leq i \leq n$$

  with the Z_i's from an $AR(1)$ time series.
- It is easy to see that

  $$E(\overline{W}) = \mu$$

  BUT, generally

  $$\mathrm{Var}(\overline{W}) > \frac{\mathrm{Var}(Z_1)}{n}$$

  how much bigger depends on ρ.

Misleading inference ignoring autocorrelation

- Just as in weighted least squares, ignoring the autocorrelation yields misleading `Std. Error` values.
- Below, we show that ignoring autocorrelation will yield incorrect confidence intervals.
  - The red curve is (an estimate of) the true density of the sample mean, while the blue curve is what we think it should be if the errors were independent.
  - The blue curve is way too optimistic.

Misleading inference ignoring autocorrelation

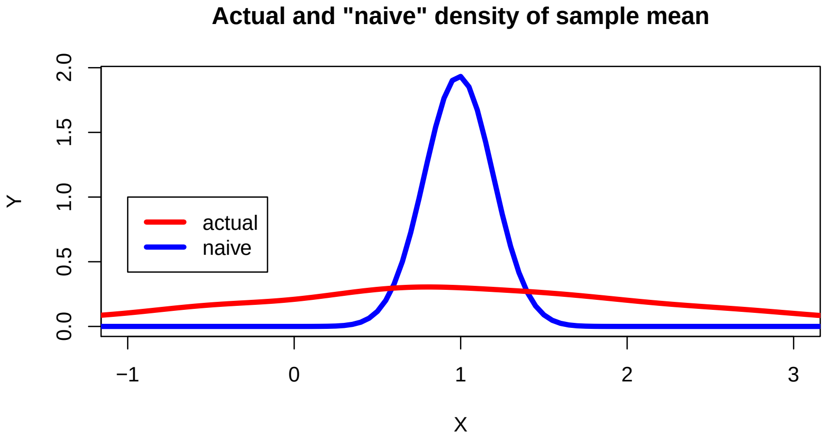
```r
ntrial = 1000

sample.mean = numeric(ntrial)
sample.var = numeric(ntrial)

for (i in 1:ntrial) {
  cur.sample = arima.sim(list(ar=rho),
    nsample) + mu
  sample.mean[i] = mean(cur.sample)
  sample.var[i] = var(cur.sample)
}

data.frame(mean=mean(sample.mean),
  sd=sqrt(mean(sample.var)))
```

```
##        mean       sd
## 1 0.9887633 2.915206
```

Misleading inference ignoring autocorrelation

```
xval = seq(-5, 5, 0.05)
Y = c(density(sample.mean)$y, dnorm(xval,
  mean = mean(sample.mean),
  sd=sqrt(mean(sample.var)/nsample)))
X = c(density(sample.mean)$x, xval)
plot(X, Y, type='n',
  main='Actual and "naive" density of sample mean',
  xlim=c(-1,3))
lines(xval,
  dnorm(xval, mean=mean(sample.mean),
    sd=sqrt(mean(sample.var) / nsample)),
  lwd=4, col='blue')
lines(density(sample.mean), lwd=4, col='red')
legend(-1,1, c('actual', 'naive'),
  col=c('red', 'blue'), lwd=rep(4,3))
```

Misleading inference ignoring autocorrelation



Regression model with auto-correlated errors (AR(1))

- Observations:

$$Y_t = \beta_0 + \sum_{j=1}^{p} X_{tj}\beta_j + \varepsilon_t, \qquad 1 \leq t \leq n$$

- Errors:

$$\varepsilon_t = \rho \cdot \varepsilon_{t-1} + \omega_t, \qquad -1 < \rho < 1$$

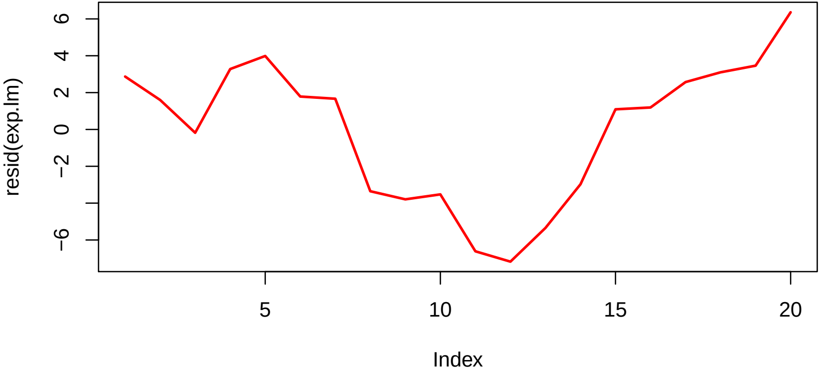
- Question: how do we determine if autocorrelation is present?
- Question: what do we do to correct for autocorrelation?

Graphical checks for autocorrelation

- A plot of residuals vs. time is helpful.
- Residuals clustered above and below 0 line can indicate autocorrelation.

Expenditure vs. stock: residuals

```
exp.lm = lm(Expenditure ~ Stock,
  data = expenditure.table)
plot(resid(exp.lm), type='l',
  lwd=2, col='red')
```



Durbin-Watson test

- In regression setting, if noise is AR(1), a simple estimate of ρ is obtained by (essentially) regressing e_t onto e_{t-1}

$$\widehat{\rho} = \frac{\sum_{t=2}^{n} (e_t e_{t-1})}{\sum_{t=1}^{n} e_t^2}.$$

- To formally test $H_0 : \rho = 0$ (i.e. whether residuals are independent vs. they are AR(1)), use Durbin-Watson test, based on

$$d \approx 2(1 - \widehat{\rho}).$$

Correcting for AR(1)

- Suppose we know ρ, we can then "whiten" the data and regressors
$$\tilde{Y}_{t+1} = Y_{t+1} - \rho Y_t, t > 1$$
$$\tilde{X}_{(t+1),j} = X_{(t+1),j} - \rho X_{t,j}, i > 1$$
for $1 \leq t \leq n-1$. This model satisfies "usual" assumptions, i.e. the errors
$$\tilde{\varepsilon}_t = \omega_{t+1} = \varepsilon_{t+1} - \rho \cdot \varepsilon_t$$
are independent $N(0, \sigma^2)$.
- For coefficients in new model $\tilde{\beta}$, $\beta_0 = \tilde{\beta}_0/(1-\rho)$, $\beta_j = \tilde{\beta}_j$.
- Problem: in general, we don't know ρ.

Two-stage regression

- As in weighted least squares, we will use a two-stage procedure.
  - Step 1: Fit linear model to unwhitened data (OLS: ordinary least squares, i.e. no prewhitening).
  - Step 2: Estimate ρ with $\widehat{\rho}$.
  - Step 3: Pre-whiten data using $\widehat{\rho}$ – refit the model.

Whitening

- Our solution in the weighted least squares and auto-correlated errors examples were the same. This procedure is generally called *whitening*.
- Consider a model

$$Y = X\beta + \epsilon, \qquad \epsilon \sim N(0, \Sigma).$$

- If Σ is invertible, then we can find a inverse square root of Σ:

$$\Sigma^{-1/2}\Sigma(\Sigma^{-1/2})^T = I, \qquad (\Sigma^{-1/2})^T\Sigma^{-1/2} = \Sigma^{-1}.$$

- Define

$$\tilde{Y} = \Sigma^{-1/2}Y, \qquad \tilde{X} = \Sigma^{-1/2}X.$$

- Then

$$\tilde{Y} = \tilde{X}\beta + \tilde{\epsilon}, \qquad \tilde{\epsilon} \sim N(0, I).$$

Generalized least squares

- The OLS estimator based on $(\tilde{Y}, \tilde{X})$ is

$$\hat{\beta}_\Sigma = (X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}Y \sim N(\beta, (X^T\Sigma^{-1}X)^{-1})$$

- It is often called the *GLS (Generalized Least Squares)* estimate based on the covariance matrix Σ.
- The OLS estimator based on (Y, X) has the sandwich form again:

$$\hat{\beta} = (X^TX)^{-1}X^TY \sim N(\beta, (X^TX)^{-1}X^T\Sigma X(X^TX)^{-1}).$$

- As in WLS, the GLS estimator with $\Sigma = \mathrm{Var}(Y)$ will generally be a more efficient estimator.
- WLS is special case when Σ is diagonal.

Interpreting results of two-stage fit

- Basically, interpretation is unchanged, but the exact degrees of freedom in the error is not exactly clear.
- Commonly applied argument:"this works for large degrees of freedom, so we hope we have enough degrees of freedom so this point is not important."
- Can treat t-statistics as Z-statistics, F's as χ^2, appealing to asymptotics:
  - t_ν, with ν large is like $N(0, 1)$;
  - $F_{j,\nu}$, with ν large is like χ^2_j/j.

Expenditure vs. stock: Durbin-Watson

```
library(car) # durbin.watson is in the "car" package
durbinWatsonTest(exp.lm)
```

```
##  lag Autocorrelation D-W Statistic p-value
##    1       0.7506122     0.3282113       0
##  Alternative hypothesis: rho != 0
```

```
rho.hat = durbinWatsonTest(exp.lm)$r
```

- Given the value of ρ, we can apply our whitening procedure.

Expenditure vs. stock

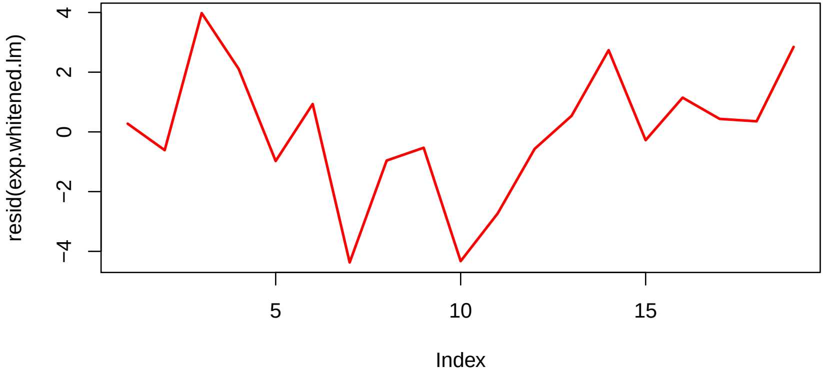
- Whitening

```
wExp = numeric(length(expenditure.table$ Expenditure) - 1)
wStock = numeric(length(expenditure.table$Expenditure) - 1)
for (i in 2:length(expenditure.table$Expenditure)) {
  wExp[i-1] = expenditure.table$Expenditure[i] -
    rho.hat * expenditure.table$Expenditure[i-1]
  wStock[i-1] = expenditure.table$Stock[i] -
    rho.hat * expenditure.table$Stock[i-1]
}
```

Expenditure vs. stock

- After whitening, we refit the model.

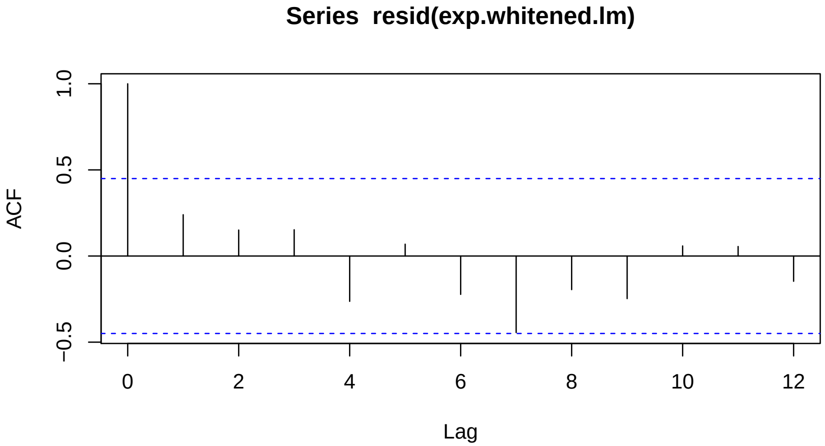
```
exp.whitened.lm = lm(wExp ~ wStock)
plot(resid(exp.whitened.lm), type='l',
  lwd=2, col='red')
```



Expenditure vs. stock

- Lastly, let's take a look at the residuals of the whitened data.
- If our whitening has been successful, this should just be a spike at 0.

```
acf(resid(exp.whitened.lm))
```



Expenditure vs. stock

- Comparing to our original fit, we see that our t statistic has changed by a factor of roughly 2.5 from 20 to 8.6!

```
Call:
lm(formula = Expenditure ~ Stock, data = expenditure.table)

Residuals:
   Min     1Q Median     3Q    Max
-7.176 -3.396  1.396  2.928  6.361

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -154.7192    19.8500  -7.794 3.54e-07 ***
Stock          2.3004     0.1146  20.080 8.99e-14 ***
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.983 on 18 degrees of freedom
Multiple R-squared:  0.9573,    Adjusted R-squared:  0.9549
F-statistic: 403.2 on 1 and 18 DF,  p-value: 8.988e-14
```

```
Call:
lm(formula = wExp ~ wStock)

Residuals:
    Min      1Q  Median      3Q     Max
-4.3737 -0.7856  0.2747  1.0408  3.9786

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -53.6959    13.6164  -3.943  0.00105 **
wStock        2.6434     0.3069   8.614 1.32e-07 ***
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.263 on 17 degrees of freedom
Multiple R-squared:  0.8136,    Adjusted R-squared:  0.8026
F-statistic:  74.2 on 1 and 17 DF,  p-value: 1.315e-07
```

Reference

- **CH** Chapter 8
- Lecture notes of Jonathan Taylor .