

Homework 7, MATH 455: Due Mon, 04/30/2018

Alexander Van Roijen

April 30, 2018

Instructions: The homework assignment editing this L^AT_EX document. Download the L^AT_EX source from the class web page and study it to learn more about L^AT_EX. Replace the text with appropriate information. Run “pdflatex” on this document.

You will submit this assignment in two parts:

1. Print out the PDF file and bring it to class, and
2. Send an e-mail to:

gang@math.binghamton.edu

before class on the due date with two attachments:

- The L^AT_EX source file, and
- The generated PDF document.

Please complete the following:

1. Finish R exercises 11.1, 11.2, 11.3, 11.4, 11.6 of the textbook. Submit your answers for **ALL** questions.

(a) 11.1

(b) 11.2

(c) 11.3

(d) 11.4

(e) 11.6

2. Finish R exercises 13.2, 13.3 of the textbook. Submit your answers for **ALL** questions.

(a) 13.2

(b) 13.3

3. Finish R exercises 8.1, 8.2, 8.6, of the textbook. Submit your answers for **ALL** questions.

(a) 8.1

i. a

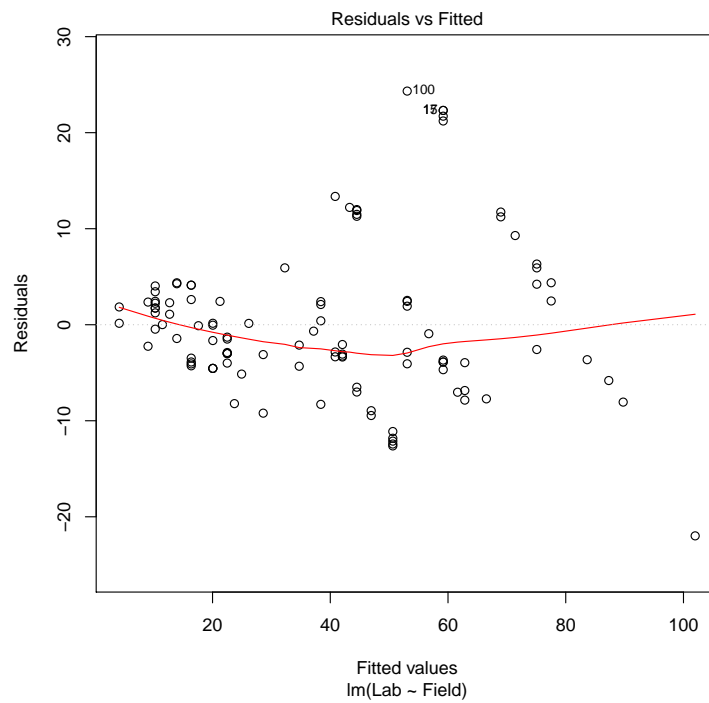


Figure 1: variance check on pipeline data

clearly there is some fanning here

ii. b

```
> summary(pipwlm)
```

Call:

```
lm(formula = Lab ~ Field, data = pipeline, weights = 1/((Field)^a_1))
```

Weighted Residuals:

Min	1Q	Median	3Q	Max
-1.7450	-0.6789	-0.2672	0.5205	2.8847

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-1.49436	0.90707	-1.647	0.102

```
Field          1.20828    0.03488  34.637   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.9795 on 105 degrees of freedom
Multiple R-squared:  0.9195, Adjusted R-squared:  0.9188
F-statistic: 1200 on 1 and 105 DF,  p-value: < 2.2e-16
```

we see some improved R squared values as we diminish the values in order to try and prevent the fanning effect

iii. c

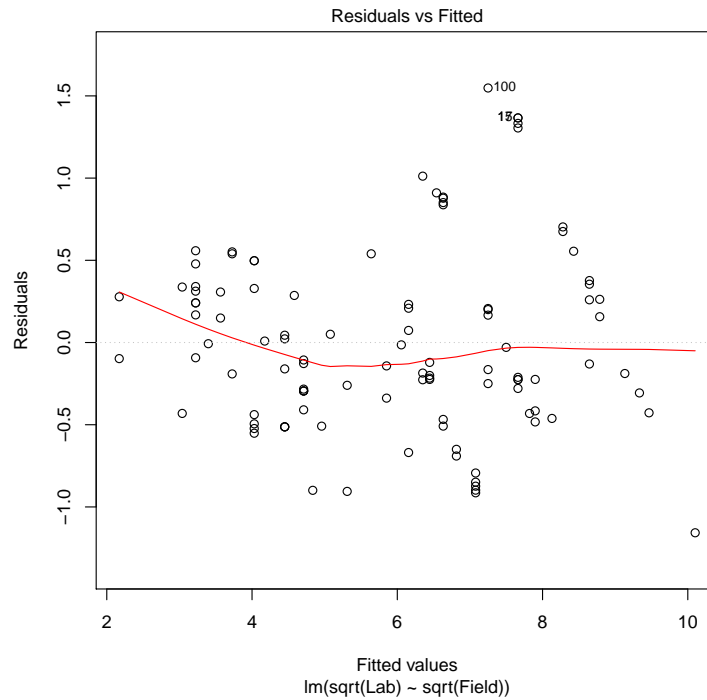


Figure 2: variance check on pipeline data after transform

This si the results of taking the square root on both the response and explanatory variables. It worked quite well.

(b) 8.2

i. a

we can see there is a correlation over time between the residuals/errors

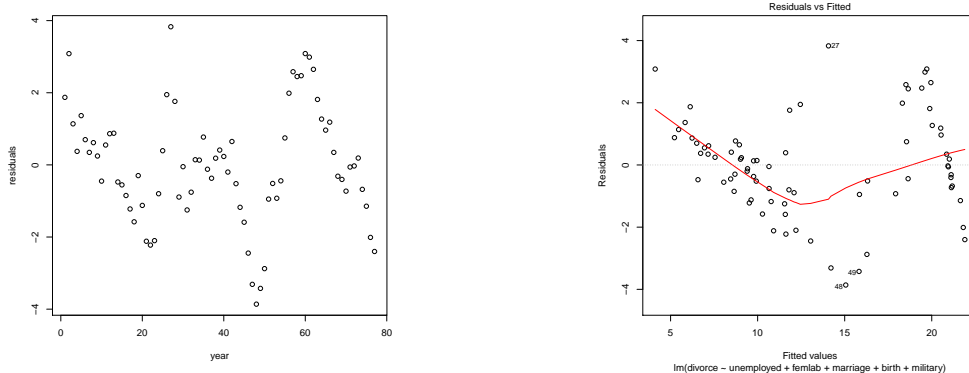


Figure 3: looking at the error correlation of Figure 4: Another look at error correlation of
divusa divusa

ii. b

```
> summary(glusalm)

Generalized least squares fit by maximum likelihood

Model: divorce ~ unemployed + femlab + marriage + birth + military
Data: divusa
AIC      BIC    logLik
179.9523 198.7027 -81.97613

Correlation Structure: AR(1)
Formula: ~year
Parameter estimate(s):
Phi
0.9715486

Coefficients:
Value Std.Error  t-value p-value
(Intercept) -7.059682  5.547193 -1.272658  0.2073
unemployed   0.107643  0.045915  2.344395  0.0219
femlab        0.312085  0.095151  3.279878  0.0016
```

marriage	0.164326	0.022897	7.176766	0.0000
birth	-0.049909	0.022012	-2.267345	0.0264
military	0.017946	0.014271	1.257544	0.2127

Correlation:

(Intr) unmply femlab marrig birth

unemployed -0.420

femlab -0.802 0.240

marriage -0.516 0.607 0.307

birth -0.379 0.041 0.066 -0.094

military -0.036 0.436 -0.311 0.530 0.128

Standardized residuals:

Min	Q1	Med	Q3	Max
-1.4509327	-0.9760939	-0.6164694	1.1375377	2.1593261

Residual standard error: 2.907665

Degrees of freedom: 77 total; 71 residual

> intervals(glusalm,which="var-cov")

Approximate 95% confidence intervals

Correlation structure:

lower	est.	upper
-------	------	-------

Phi	0.6528097	0.9715486	0.9980192
-----	-----------	-----------	-----------

attr("label")

[1] "Correlation structure:"

Residual standard error:

lower	est.	upper
-------	------	-------

0.7974404	2.9076645	10.6020628
-----------	-----------	------------

we can see that unemployed has become significant, in the previous model, the pvalue was higher.

Further their correlation is significant, we see a positive correlation with a confidence interval that is quite strong

- iii. c Personally, I believe these are correlated over the years mainly due to the warts the data set covers. Baby boomers are all likely to get married around the same time, and thus divorce in similar times as well. Further, War usually causes couples to get married just before leaving for service or after. Thus when they return they will realize they werent meant to be and similarly get divorced at similar times.

(c) 8.6

- i. a

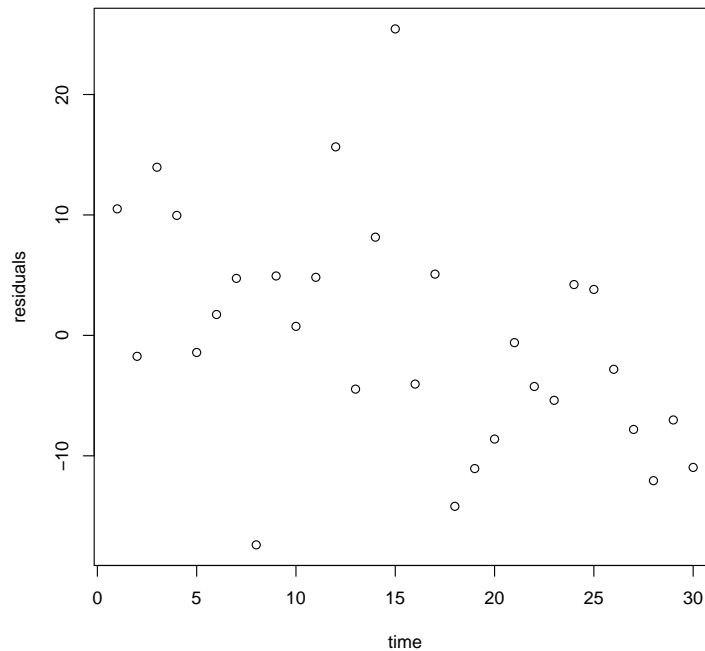


Figure 5: we can see a somewhat linear trend over time that is decreasing.

Not a strong indicator, but something

- ii. b

Generalized least squares fit by REML

Model: taste ~ . - time

Data: c2

AIC BIC logLik

214.94 222.4886 -101.47

Correlation Structure: AR(1)

Formula: ~time

Parameter estimate(s):

Phi

0.2641944

Coefficients:

Value Std.Error t-value p-value

(Intercept) -30.332472 20.273077 -1.496195 0.1466

Acetic 1.436411 4.876581 0.294553 0.7707

H2S 4.058880 1.314283 3.088284 0.0047

Lactic 15.826468 9.235404 1.713674 0.0985

Correlation:

(Intr) Acetic H2S

Acetic -0.899

H2S 0.424 -0.395

Lactic 0.063 -0.416 -0.435

Standardized residuals:

Min Q1 Med Q3 Max

-1.64546468 -0.63861716 -0.06641714 0.52255676 2.41323021

Residual standard error: 10.33276

Degrees of freedom: 30 total; 26 residual


```
> intervals(cgls,which="var-cov")
Approximate 95% confidence intervals
```

```
Correlation structure:
lower      est.      upper
Phi -0.1690265 0.2641944 0.6118599
attr(,"label")
[1] "Correlation structure:"
```

```
Residual standard error:
lower      est.      upper
7.62646 10.33276 13.99940
```

We can see that the confidence interval include 0, and thus we can not really trust this correlation.

```
iii. > clm2 = lm(taste~.,c2)
> summary(clm2)
```

```
Call:
lm(formula = taste ~ ., data = c2)
```

```
Residuals:
Min      1Q  Median      3Q     Max
-22.3523  -4.9735  -0.5089   4.8531  23.1311
```

```
Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) -36.6127    17.9845  -2.036  0.05250 .
Acetic        4.1275     4.2556   0.970  0.34139
H2S           3.5387     1.1315   3.127  0.00444 **
Lactic       17.9527     7.7875   2.305  0.02973 *
```

```
time          -0.5459      0.2043  -2.672  0.01306 *
```

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 9.112 on 25 degrees of freedom
```

```
Multiple R-squared:  0.7291, Adjusted R-squared:  0.6858
```

```
F-statistic: 16.83 on 4 and 25 DF,  p-value: 8.205e-07
```

Unlike the GLS, our OLS thinks time is significant! Very funny. However, this is not contradictory, LS and GLS are quite different. This is explained in the next part.

iv. d

in the GLS, we are looking at how correlated the error or noise is over "time", or consecutive entries unlike our ordinary LS. Within the OLS the time value is being included to see how it may provide information on our response. The difference lies within the relations. In OLS it changes the significance and value based on a linear combination within each entry. In residuals, we are only considering the impact of the time variable **AFTER** the coefficients have been established

v. e

if i was told that the entries were not in chronological order, then this would make it purely coincidental that consecutive entries are related, and we should randomize their order to avoid the seemingly correlated entries