STARTED 4/18/22

Monday (Week 13, lecture 33)

STARTED by going over MLS Anemia.R

@ 25 min mark

START Logistic.r

Go over Michelin.r

START Wednesday 4/20/22 (week 13/ lecture 34)

Stat 608 Chapter 9

*STARTED by going over Ex 8.5.r

* If marginal model plot for x doesn't look good:

Serially Correlated Errors

* Note: for variable X, draw density plots for y=1 & y=0. If they look normally distributed, but w/ different variances ⇒ add X^2 to model. If the distributions look right skewed ⇒ add log(x) to model.

START friday 4/22/22 (week 13, lecture 35)
@ 23 min mark (went over HW6 sols first)

+ Motivation

- In many situations, data are collected over time
- Measurements at nearby time points expected to be positively correlated
- This correlation violates the assumption of independent errors that we had with linear regression models *
- Autocorrelation is a commonly-used measure of correlation between time points
- Generalized Least Squares (GLS) can be used to fit models with correlation over time *
- We can transform GLS models to LS models in order to access the usual diagnostic plots

↳ * b/c our usual residual plots to diagnose model validity don't work in the presence of autocorrelation.

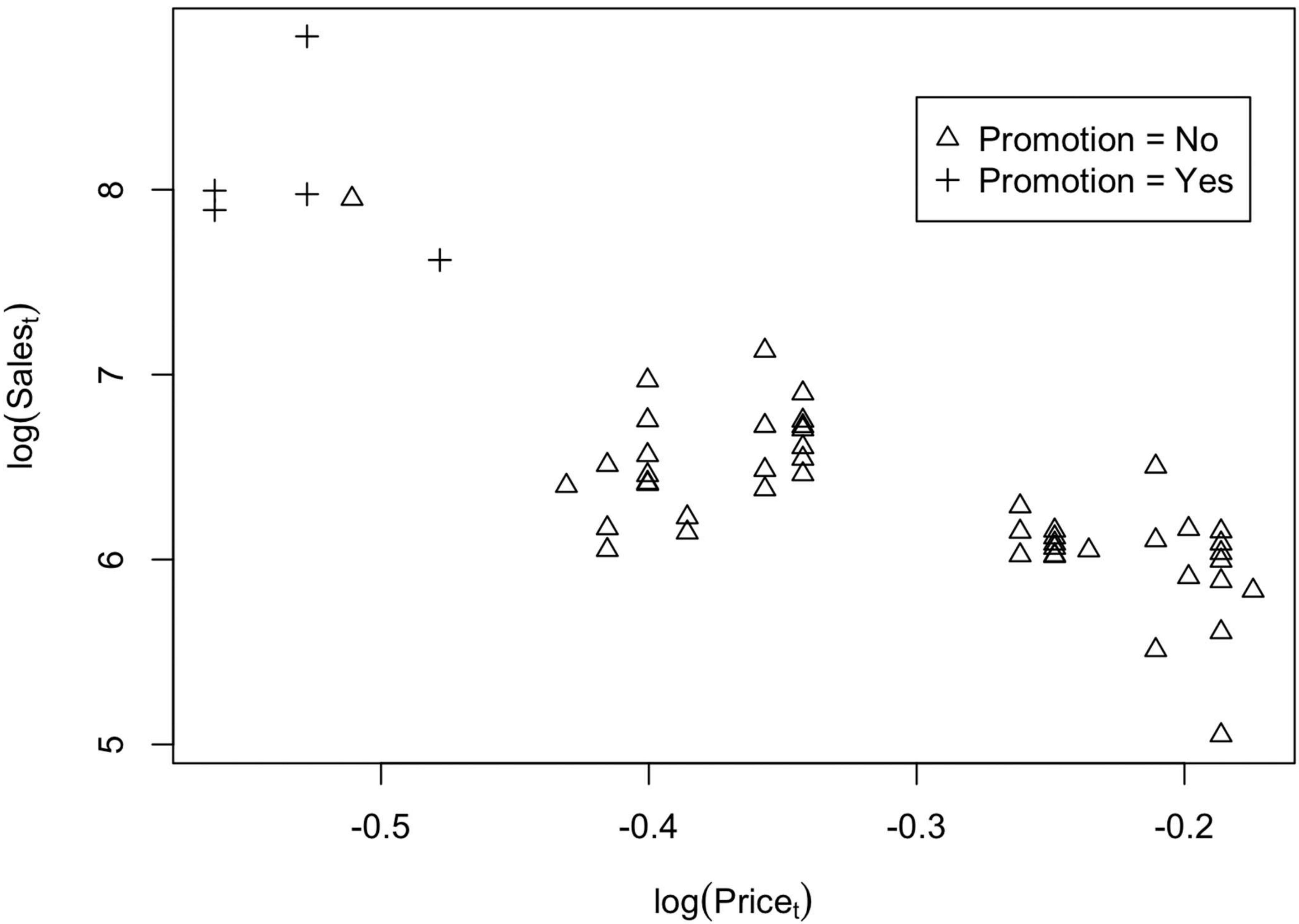
Autocorrelation

Example: Price Elasticity of Food Product

- Goal: Understand the effect of price on sales
  - Data collected weekly over a year
- Original model we considered earlier:

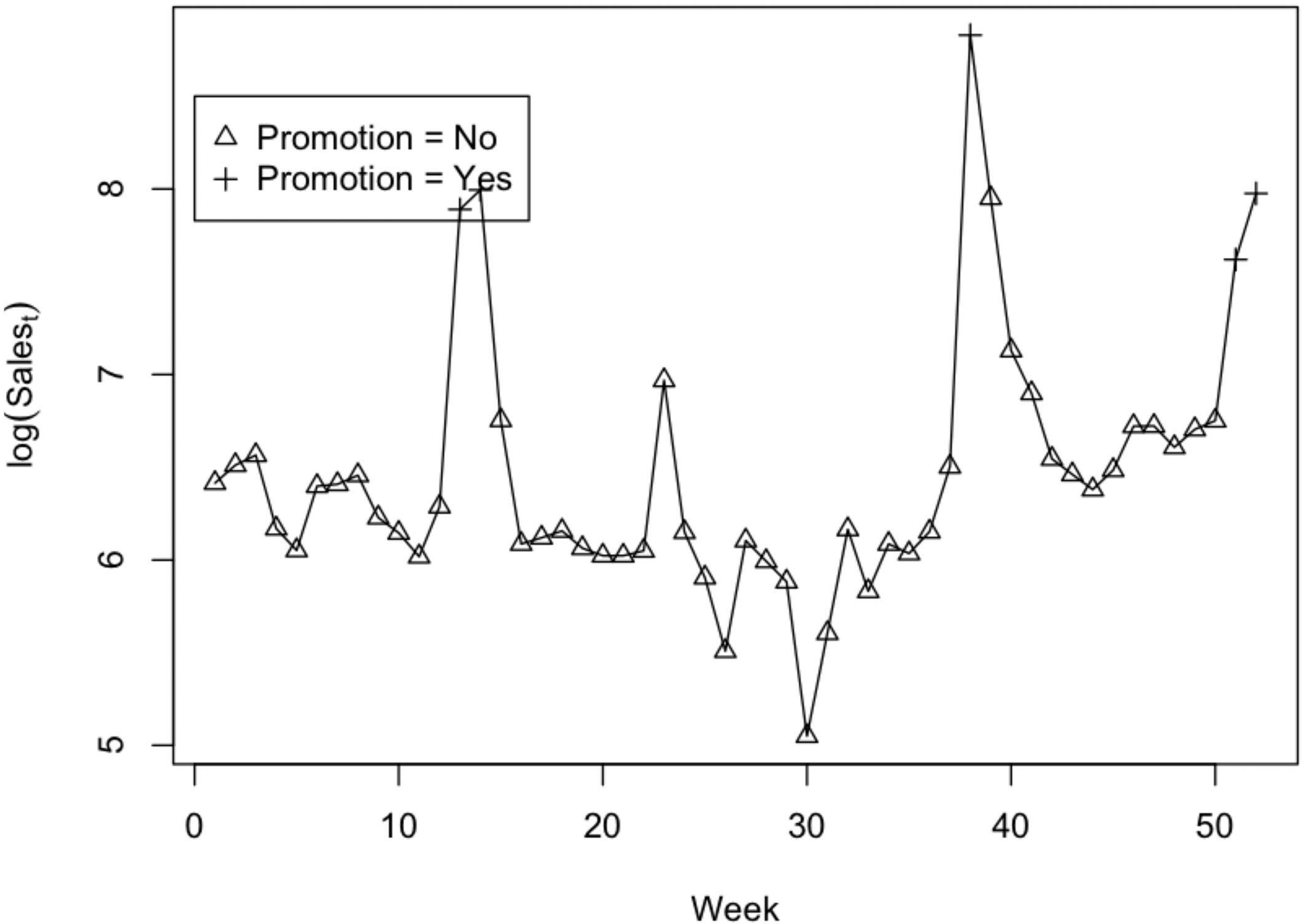
$$\log(\text{Sales}_t) = \beta_0 + \beta_1 \log(\text{Price}_t) + e$$

- Other available predictor variables:
  - Week = week of the year
  - Promotion_t = dummy variable indicating whether a promotion occurred in week *t*:
    - 0 = no promotion
    - 1 = price reduction

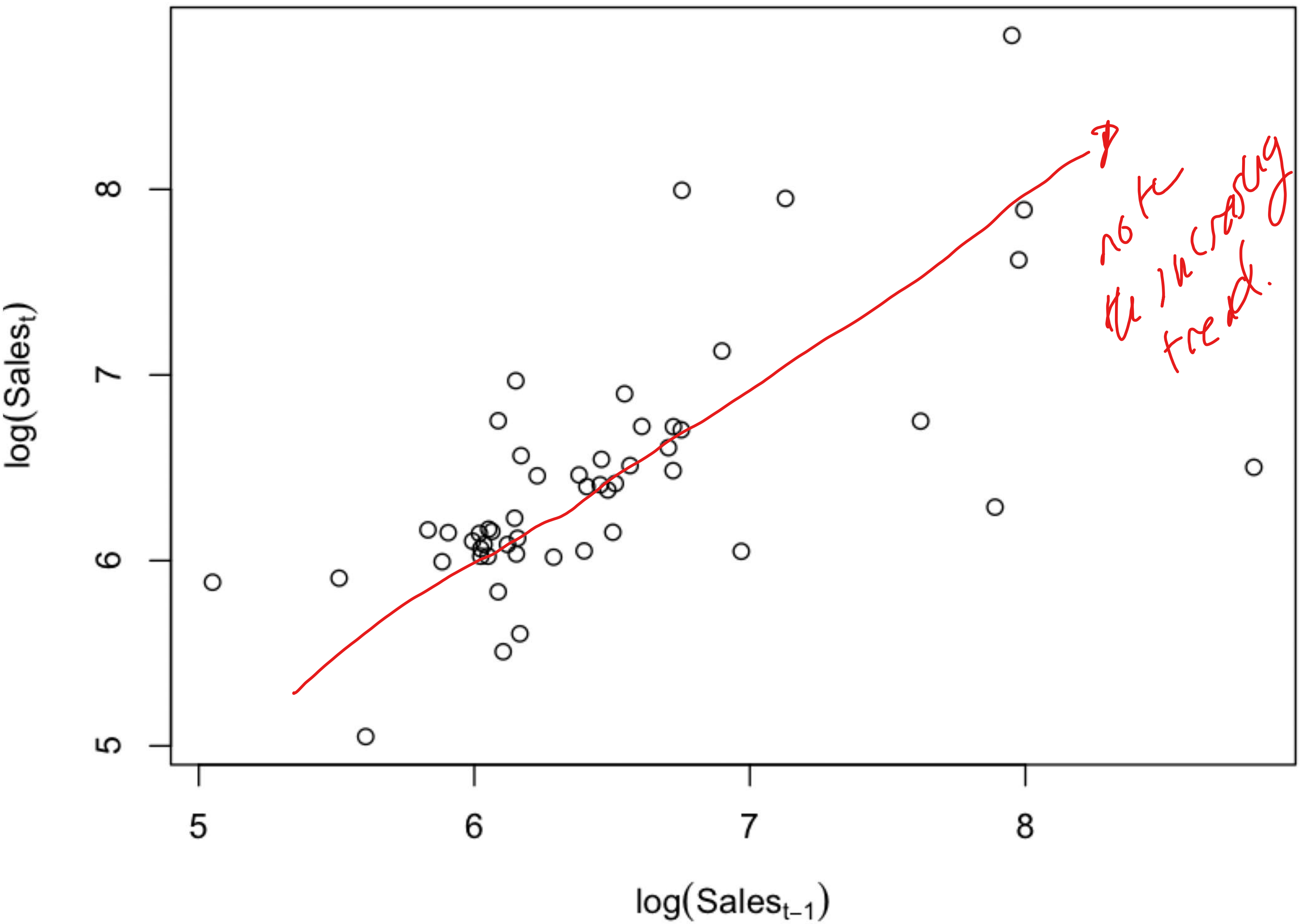


Promotion = No
Promotion = Yes
log(Sales$_t$)
log(Price$_t$)
5
6
7
8
-0.5
-0.4
-0.3
-0.2

”Time Series Plot”



“Lag Plot”



Comments

- Promotion has a big effect on Sales
- Positive correlation:
  - Weeks with above average Sales followed generally by above average Sales
  - Weeks with below average Sales followed generally by below average Sales
- Sales in Week *t* apparently positively correlated with Sales in Week *t*-1.

Autocorrelation

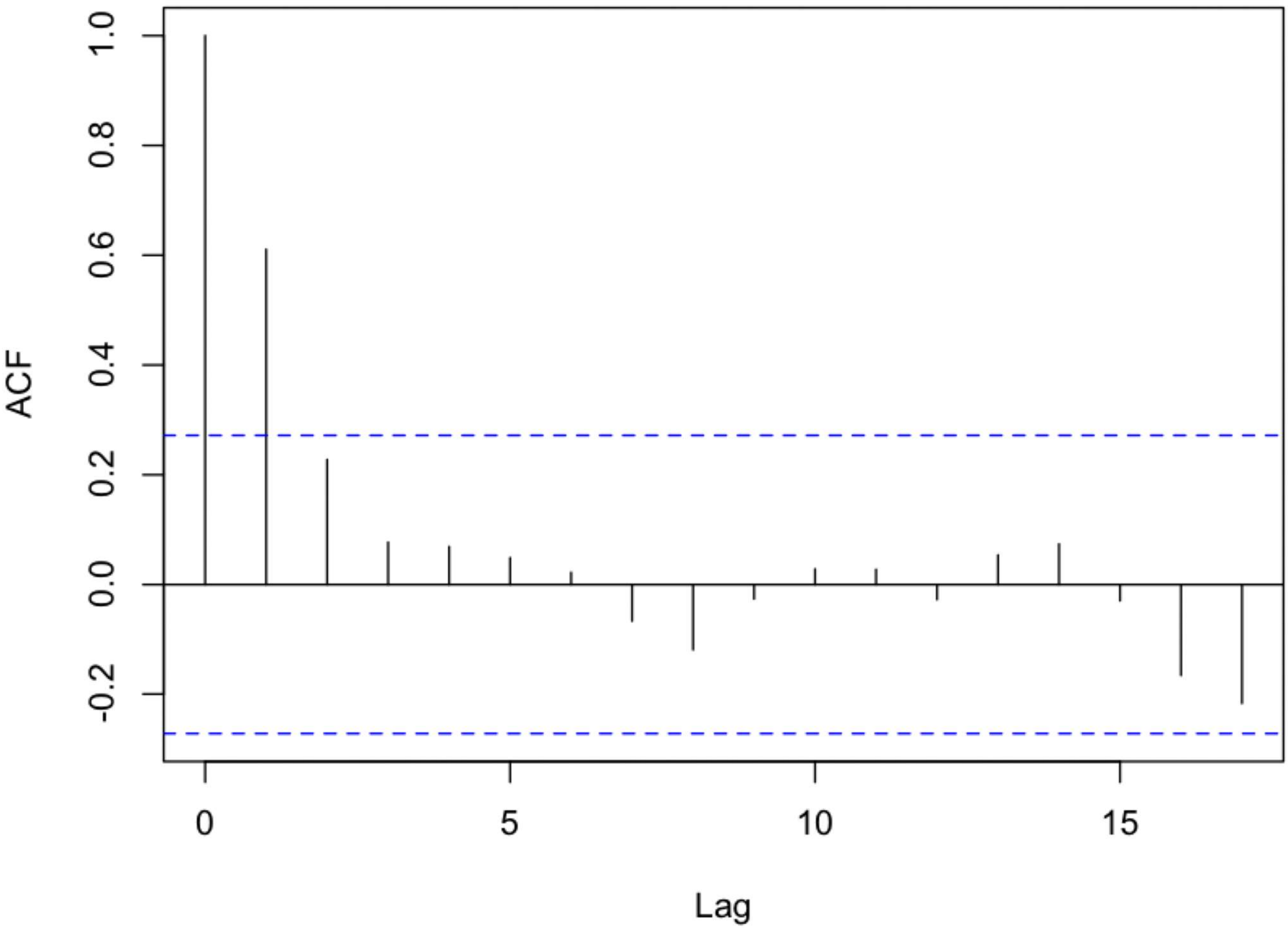
- What about Sales in Week *t* vs. Sales in Weeks *t*-2, *t*-3, etc.?

$$\text{Autocorrelation}(l) = \frac{\sum_{t=l+1}^{n} (y_t - \bar{y})(y_{t-l} - \bar{y})}{\sum_{t=1}^{n} (y_t - \bar{y})^2}$$

  - Called the autocorrelation of lag *l*
  - The correlation between *Y* and values of *Y* lagged by *l* periods
  - Autocorrelations bigger than $2/\sqrt{n}$ or smaller than $-2/\sqrt{n}$ statistically significantly different from 0

+

"Autocorrelation Plot"

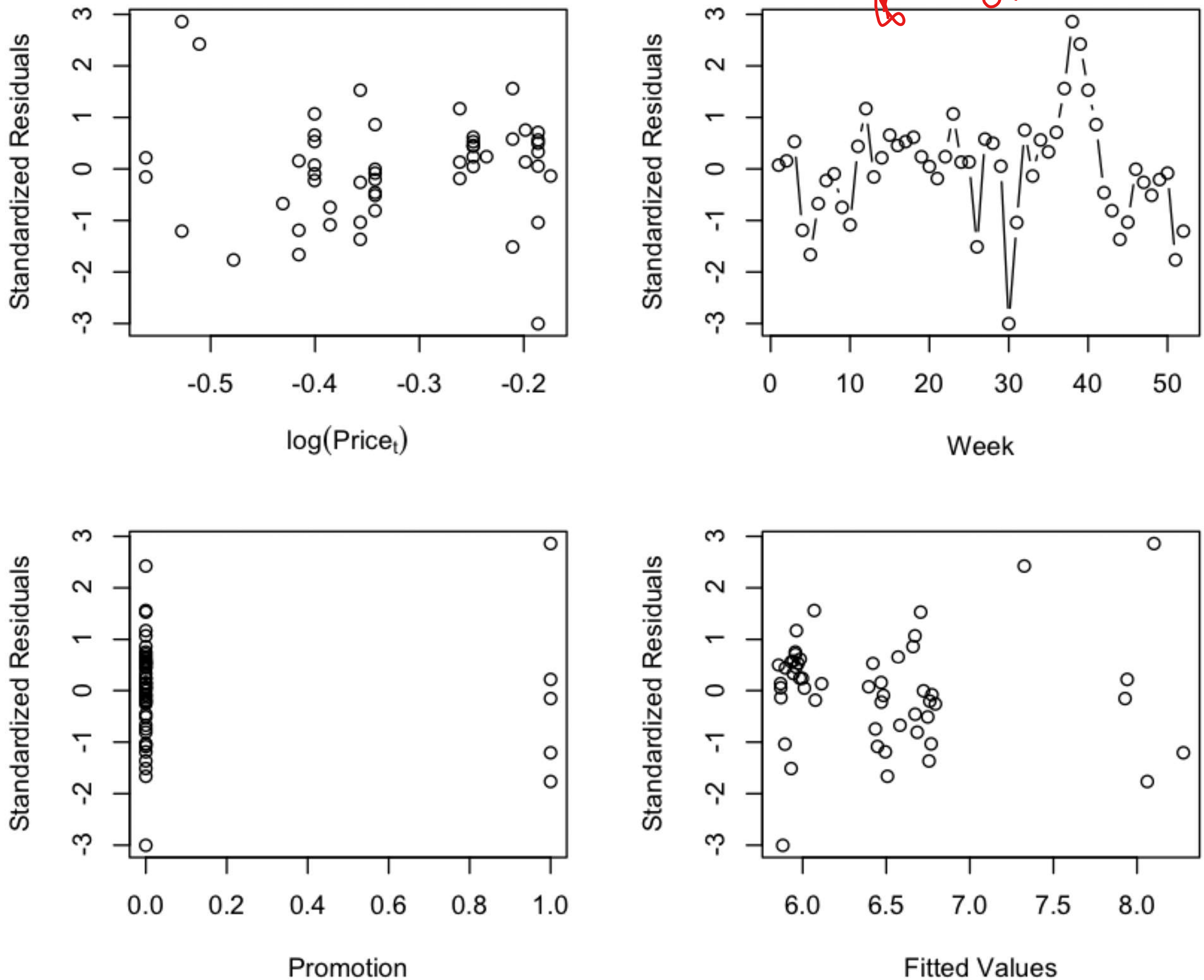


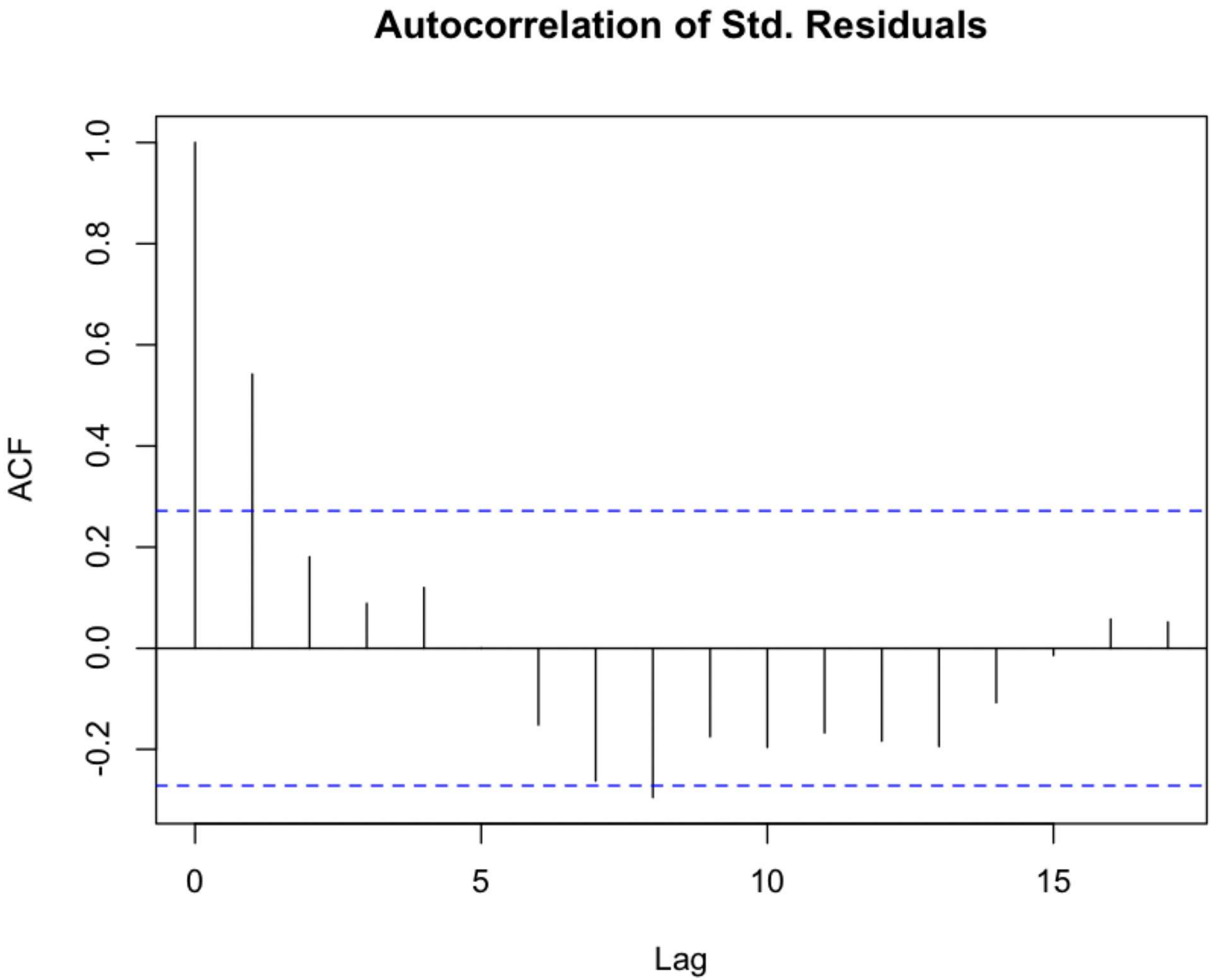
Ignoring the autocorrelation

- Let's first consider a model that ignores the autocorrelation:

$$\log\left(\text{Sales}_t\right) = \beta_0 + \beta_1 \log\left(\text{Price}_t\right) + \beta_2 t + \beta_3 \text{Promotion}_t + e$$

  - Assuming (naively) that the errors are independent see evidence of autocorrelation




Autocorrelation of Std. Residuals
ACF
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
0
5
10
15
Lag

Move to chapter9.r code.

STOP Friday 4/22/22 (week 13, lecture 35)

START Monday 4/25/22 (week 14, lecture 36)

+

Generalized Least Squares

GLS when errors are AR(1)

- Simple starting model:

$$Y_t = \beta_0 + \beta_1 x_t + e_t, \text{ where } e_t = \rho e_{t-1} + \nu_t \text{ and } \nu_t \text{ are iid } N\left(0, \sigma_\nu^2\right)$$

- We have:

$$\mathrm{E}\left(e_t\right) = \mathrm{E}\left(\rho e_{t-1} + \nu_t\right) = \rho \mathrm{E}\left(e_{t-1}\right) + \mathrm{E}\left(\nu_t\right) = 0$$

- And:

$$E[(e_t - \underbrace{E[e_t]}_{\to 0})^2] = E[e_t^2]$$

$$E[e_{t-1}^2] = var(e_{t-1}) = \sigma_e^2$$

$$E[\nu_t^2] = var(\nu_t^2) = \sigma_\nu^2$$

$$\begin{aligned} \sigma_e^2 &= \operatorname{Var}\left(e_t\right) = \mathrm{E}\left(e_t^2\right) \\ &= \mathrm{E}\left[\left(\rho e_{t-1} + \nu_t\right)^2\right] \\ &= \rho^2 \underline{\mathrm{E}\left(e_{t-1}^2\right)} + \mathrm{E}\left(\nu_t^2\right) + 2\rho \mathrm{E}\left(\cancel{e_{t-1}}\right)_{\to 0} \mathrm{E}\left(\cancel{\nu_t}\right)_{\to 0} \\ \sigma_e^2 &= \rho^2 \sigma_e^2 + \sigma_\nu^2 \iff \sigma_e^2 - \rho^2\sigma_e^2 = \sigma_\nu^2 \iff \sigma_e^2 = \frac{\sigma_\nu^2}{1-\rho^2} \end{aligned}$$

GLS when errors are AR(1)

- Rearranging last term:

$$\sigma_e^2 = \frac{\sigma_\nu^2}{1-\rho^2}$$

- Thus:

$$\mathrm{Corr}\left(e_t, e_{t-1}\right) = \frac{\mathrm{Cov}\left(e_t, e_{t-1}\right)}{\sqrt{\mathrm{Var}\left(e_t\right)\mathrm{Var}\left(e_{t-1}\right)}} = \frac{\mathrm{E}\left(e_t e_{t-1}\right)}{\sqrt{\sigma_e^2 \sigma_e^2}} = \rho$$

since

$$\mathrm{E}\left(e_t e_{t-1}\right) = \mathrm{E}\left[\left(\rho e_{t-1} + \nu_t\right) e_{t-1}\right] = \rho \mathrm{E}\left(e_{t-1}^2\right) + \mathrm{E}\left(\nu_t\right)\mathrm{E}\left(e_{t-1}\right) = \rho \sigma_e^2$$

- Similarly:

$$\mathrm{Corr}\left(e_t, e_{t-l}\right) = \rho^l, \quad l = 1, 2, \ldots$$

$$\Sigma = \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^k \\ & 1 & & & \vdots \\ & & \ddots & & \vdots \\ \text{symmetric} & & & \ddots & \vdots \\ & & & & 1 \end{bmatrix}$$

GLS when errors are AR(1)

β: we fit regression model using LS (ignoring Auto-correlation)

- We have that:

$$\mathrm{E}\left(\hat{\beta}_{1,\mathrm{LS}}\right) = \beta_1$$

- But:

$$\mathrm{Var}\left(\hat{\beta}_{1,\mathrm{LS}}\right) = \frac{\sigma_e^2}{SXX}\left(1 + \frac{1}{SXX}\sum_{i \neq j}\sum (x_i - \bar{x})(x_j - \bar{x})\,\rho^{|i-j|}\right)$$

- If the errors are independent: $\rho = 0$

(pos corr $\Rightarrow$ var($\hat{\beta}$) < it should be
$\Rightarrow$ p-value < they should be
$\Rightarrow$ CI narrower than they should be.)

$$\mathrm{Var}\left(\hat{\beta}_{1,\mathrm{LS}}\right) = \frac{\sigma_e^2}{SXX}$$

*

- Thus, if we ignore correlation, we get unbiased coefficient estimates, but the standard errors will be wrong

GLS when errors are AR(1)

- Let $\mathbf{Y}$ be an $(n \times 1)$ response vector and $\mathbf{X}$ be an $n \times (p+1)$ design matrix
- Let β be the $(p+1)$ vector of regression coefficients and $\mathbf{e}$ be the $(n \times 1)$ vector of error terms
- Our model is:

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{e}, \text{ where } \mathbf{e} \sim N(\mathbf{0}, \Sigma)$$

where Σ is a symmetric $(n \times n)$ matrix with (i, j) element equal to $\text{Cov}(e_i, e_j)$

GLS when errors are AR(1)

- Consider the AR(1) case:

$$e_t = \rho e_{t-1} + \nu_t \text{ where } \nu_t \text{ are i.i.d. } N\left(0, \sigma_\nu^2\right)$$

- It can be shown that

$$\Sigma = \sigma_e^2 \begin{pmatrix} 1 & \rho & \cdots & \rho^{n-1} \\ \rho & 1 & \cdots & \vdots \\ \vdots & & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \cdots & 1 \end{pmatrix} = \frac{\sigma_\nu^2}{1-\rho^2} \begin{pmatrix} 1 & \rho & \cdots & \rho^{n-1} \\ \rho & 1 & \cdots & \vdots \\ \vdots & & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \cdots & 1 \end{pmatrix}$$

since $\mathrm{Cov}\left(e_t, e_{t-1}\right) = \mathrm{E}\left(e_t e_{t-1}\right) = \rho \sigma_e^2$

GLS when errors are AR(1)

- The log-likelihood is

$$\log\left(L\left(\beta, \rho, \sigma_e^2 | \mathbf{Y}\right)\right)$$
$$= -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log\left(\det(\Sigma)\right) - \frac{1}{2}\left(\mathbf{Y} - \mathbf{X}\beta\right)' \Sigma^{-1}\left(\mathbf{Y} - \mathbf{X}\beta\right)$$

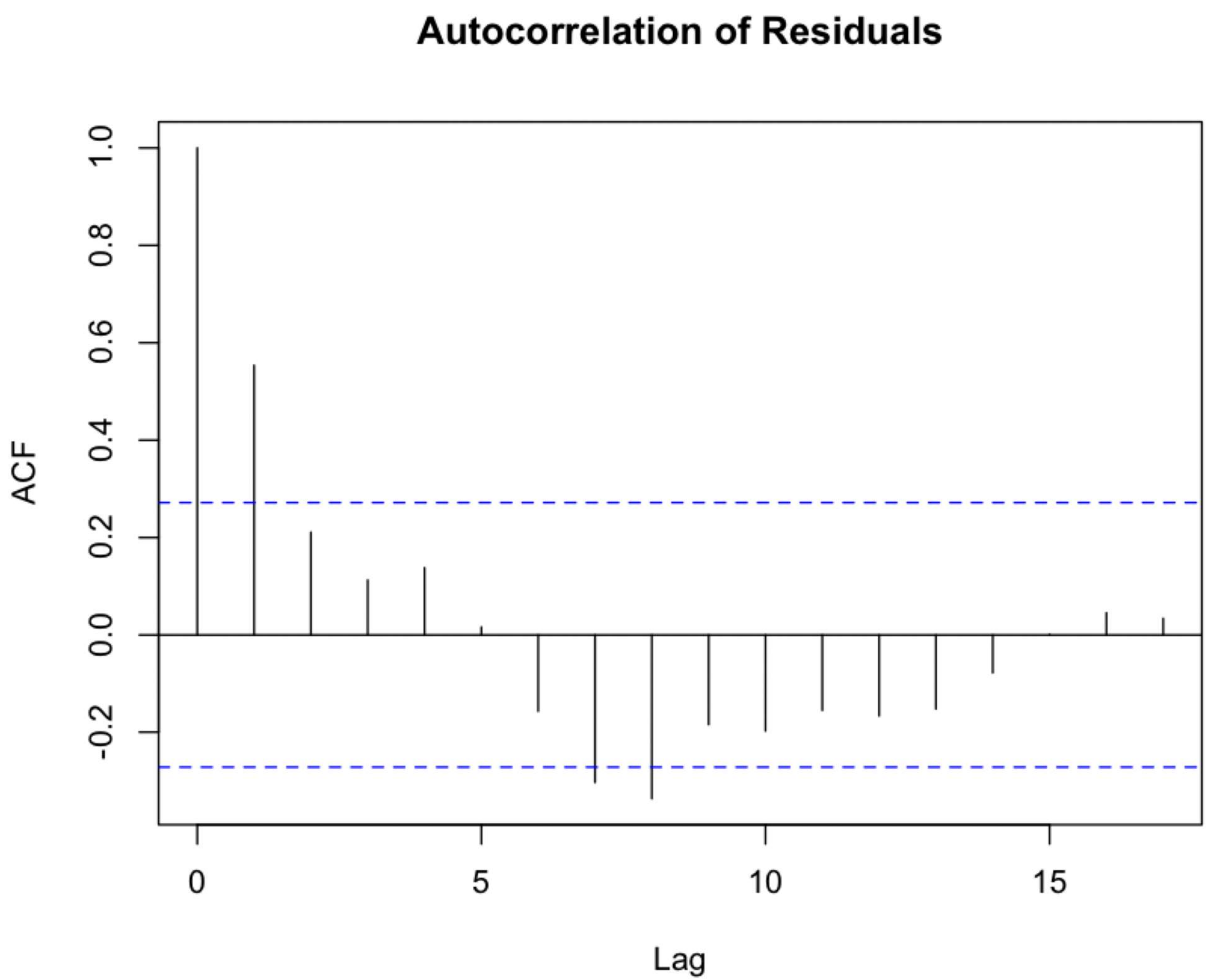
- Given estimates of ρ and σ_e^2, the GLS estimate of β is:

$$\hat{\beta}_{\mathrm{GLS}} = \left(\mathbf{X}'\hat{\Sigma}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\hat{\Sigma}^{-1}\mathbf{Y}$$

$$\hat{\beta}_{LS} = (X' \Sigma_{LS}^{-1} X)^{-1} X' \Sigma_{LS}^{-1} Y \qquad (\Sigma_{LS} = \sigma_e^2 I)$$
$$= \left(X' \frac{1}{\sigma_e^2} I X\right)^{-1} X' \frac{1}{\sigma_e^2} I Y$$
$$= \sigma_e^2 (X'X)^{-1} \left(\frac{1}{\sigma_e^2}\right) X'Y$$
$$= (X'X)^{-1} X'y$$

GLS when errors are AR(1)

```
Generalized least squares fit by maximum likelihood
  Model: log(Sales) ~ log(Price) + Promotion + Week

Correlation Structure: AR(1)
 Formula: ~Week
 Parameter estimate(s):
      Phi
0.5503593

Coefficients:
                Value Std.Error   t-value p-value
(Intercept)  4.675667 0.2383703 19.615142   0.000
log(Price)  -4.327391 0.5625564 -7.692368   0.000
Promotion    0.584650 0.1671113  3.498565   0.001
Week         0.012517 0.0046692  2.680813   0.010

Residual standard error: 0.2740294
Degrees of freedom: 52 total; 48 residual
```



Autocorrelation of Residuals
ACF
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
0
5
10
15
Lag

Transforming an AR(1) model into a model with iid errors

- We want to transform the model

$$Y_t = \beta_0 + \beta_1 x_t + e_t = \beta_0 + \beta_1 x_t + \rho e_{t-1} + \nu_t$$

into a model with uncorrelated errors so we can use least squares for diagnostics

In iid case we could do plots of std. residuals vs fitted values
" " vs covariates
and QQ plots of std residuals to check if model is valid.

Transforming an AR(1) model into a model with iid errors

- First, note that we can write:

$$\rho Y_{t-1} = \rho\beta_0 + \rho\beta_1 x_{t-1} + \rho e_{t-1}$$

- Subtracting this from the first model equation gives:

$$Y_t - \rho Y_{t-1} = \beta_0 + \beta_1 x_t + e_t - (\rho\beta_0 + \rho\beta_1 x_{t-1} + \rho e_{t-1})$$

- Since $e_t = \rho e_{t-1} + \nu_t$ we have:

$$\begin{aligned} Y_t - \rho Y_{t-1} &= \beta_0 + \beta_1 x_t + \rho e_{t-1} + \nu_t - (\rho\beta_0 + \rho\beta_1 x_{t-1} + \rho e_{t-1}) \\ &= (1-\rho)\,\beta_0 + \beta_1\,(x_t - \rho x_{t-1}) + \nu_t \end{aligned}$$

Transforming an AR(1) model into a model with iid errors

- Now make the transformation:

$$Y_t^* = Y_t - \rho Y_{t-1}, \;\; x_{t1}^* = 1 - \rho, \;\text{ and }\; x_{t2}^* = x_t - \rho x_{t-1} \;\text{ for }\; t = 2, \ldots, n$$

- The last model equation can now be rewritten:

$$Y_t^* = \beta_0 x_{t1}^* + \beta_1 x_{t2}^* + v_t, \;\text{ for }\; t = 2, \ldots, n$$

now we have iid errors after transformation

Transforming an AR(1) model into a model with iid errors

- Still need to deal with the first observation

$$Y_1 = \beta_0 + \beta_1 x_1 + e_1, \text{ where } \operatorname{Var}(e_1) = \sigma_e^2 = \frac{\sigma_\nu^2}{1-\rho^2}$$

- Multiply each term by $\sqrt{1-\rho^2}$:

$$\sqrt{1-\rho^2} Y_1 = \sqrt{1-\rho^2}\beta_0 + \sqrt{1-\rho^2}\beta_1 x_1 + \sqrt{1-\rho^2} e_1$$

- Make the transformation:

$$Y_1^* = \sqrt{1-\rho^2} Y_1, \ x_{t1}^* = \sqrt{1-\rho^2}, \text{ and } x_{t2}^* = \sqrt{1-\rho^2} x_1$$

- Then can write:

$$\operatorname{var}(e_1^*) = (1-\rho^2)\frac{\sigma_v^2}{1-\rho^2} = \sigma_v^2$$

$$Y_1^* = \beta_0 x_{11}^* + \beta_1 x_{12}^* + e_1^*$$

Transforming an AR(1) model into a model with iid errors

- We can equivalently define $Y_1^* = Y_1$ and

$$Y_t^* = (Y_t - \rho Y_{t-1})\sqrt{1-\rho^2},\ t = 2, \ldots, n$$

STOP Monday 4/25/22 (Week 14, Lecture 36)

START Wednesday 5/27/22 (week 14, lecture 37)
(started going over Assign 7 #2 until 15 min mark)

+

A General Approach to Transforming GLS to LS

- Recall that for the linear model $\mathbf{Y} = \mathbf{X}\beta + \mathbf{e}$, the GLS model estimate is:

$$\hat{\beta}_{\mathrm{GLS}} = \left(\mathbf{X}'\Sigma^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\Sigma^{-1}\mathbf{Y}$$

  where Σ is a symmetric $(n \times n)$ matrix with (i, j) element

  equal to $\mathrm{Cov}\,(e_i, e_j)$

- Since Σ is symmetric and positive definite, it can be written as:

$$\Sigma = \mathbf{S}\mathbf{S}'$$

  where $\mathbf{S}$ is a lower triangular matrix with positive diagonals

A General Approach to Transforming GLS to LS

- The matrix $\mathbf{S}$ can be thought of as the "square root" of Σ
- Multiplying each side of our model equation by $\mathbf{S}^{-1}$:

$$\mathbf{S}^{-1}\mathbf{Y} = \mathbf{S}^{-1}\mathbf{X}\beta + \mathbf{S}^{-1}\mathbf{e}$$

since $\left(\mathbf{S}^{-1}\right)' = \left(\mathbf{S}'\right)^{-1}$

- We now have that:

$$\mathrm{Var}\left(\mathbf{S}^{-1}\mathbf{e}\right) = \mathbf{S}^{-1}\mathrm{Var}\left(\mathbf{e}\right)\left(\mathbf{S}^{-1}\right)' = \mathbf{S}^{-1}\Sigma\left(\mathbf{S}^{-1}\right)' = \mathbf{S}^{-1}\mathbf{S}\mathbf{S}'\left(\mathbf{S}'\right)^{-1} = \mathbf{I}$$

I.e., our transformed model has uncorrelated errors

A General Approach to Transforming GLS to LS

- Thus, we apply the transformation:

$$\mathbf{Y}^* = \mathbf{S}^{-1}\mathbf{Y},\ \mathbf{X}^* = \mathbf{S}^{-1}\mathbf{X},\ \text{and } \mathbf{e}^* = \mathbf{S}^{-1}\mathbf{e}$$

and then we fit the linear model with uncorrelated errors:

$$\mathbf{Y}^* = \mathbf{X}^*\beta + \mathbf{e}^*$$

- We can now obtain the GLS estimate of β using least squares

A General Approach to Transforming GLS to LS

- Proving that $\hat{\beta}^*_{\text{LS}}$ for the transformed model equals $\hat{\beta}_{\text{GLS}}$:

$$\begin{aligned}
\hat{\beta}^*_{\text{LS}} = (\mathbf{X}^*\mathbf{X}^*)^{-1}\mathbf{X}^*\mathbf{Y}^* &= \left((\mathbf{S}^{-1}\mathbf{X})'(\mathbf{S}^{-1}\mathbf{X})\right)^{-1}(\mathbf{S}^{-1}\mathbf{X})'(\mathbf{S}^{-1}\overset{\mathbf{Y}}{\cancel{\mathbf{X}}}) \\
&= \left(\mathbf{X}'(\mathbf{S}^{-1})'\mathbf{S}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'(\mathbf{S}^{-1})'\mathbf{S}^{-1}\mathbf{Y} \\
&= (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}\mathbf{Y} = \hat{\beta}_{\text{GLS}}
\end{aligned}$$

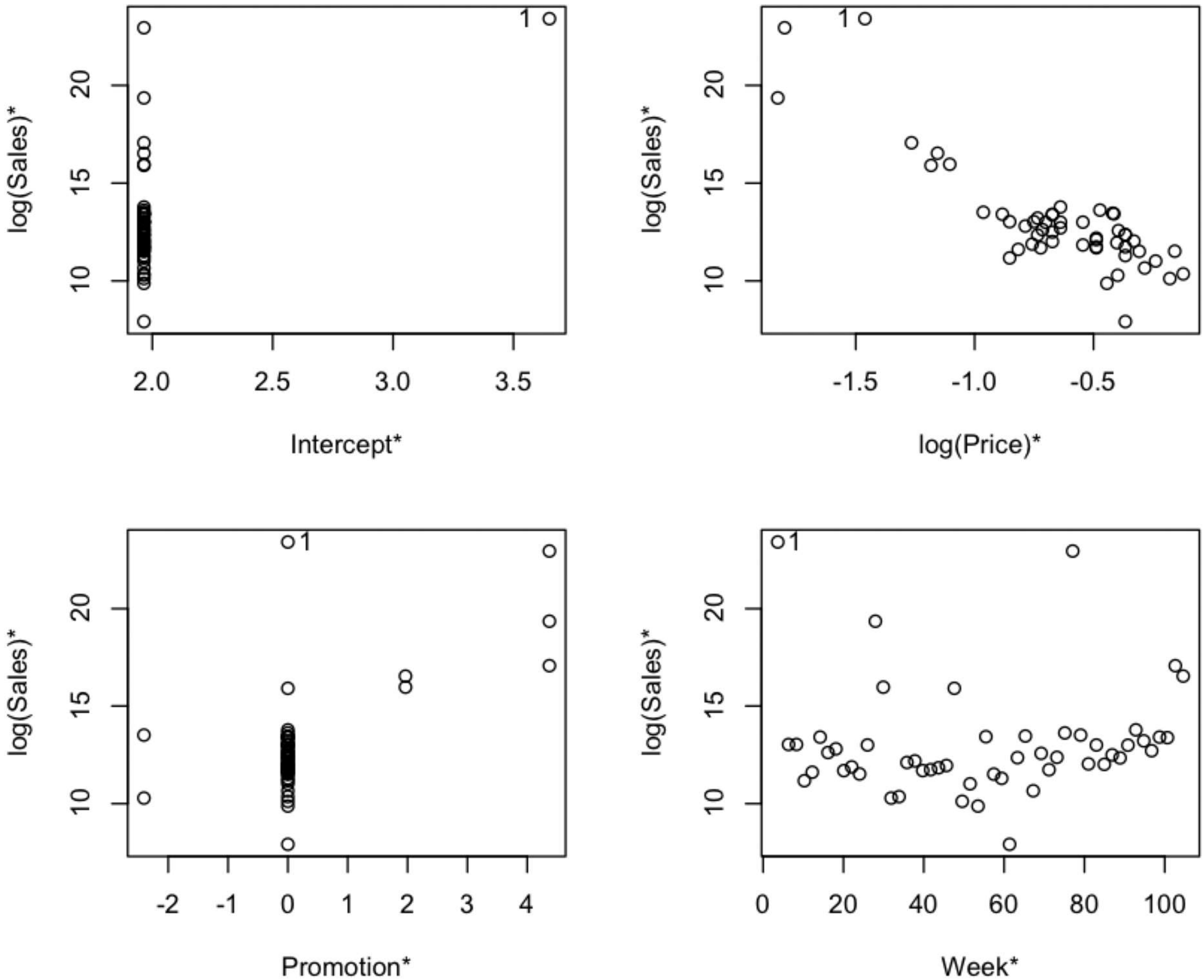
We used the fact that

$$\Sigma^{-1} = (\mathbf{S}\mathbf{S}')^{-1} = (\mathbf{S}')^{-1}\mathbf{S}^{-1} = (\mathbf{S}^{-1})'\mathbf{S}^{-1}, \text{ since } (\mathbf{A}')^{-1} = (\mathbf{A}^{-1})'$$

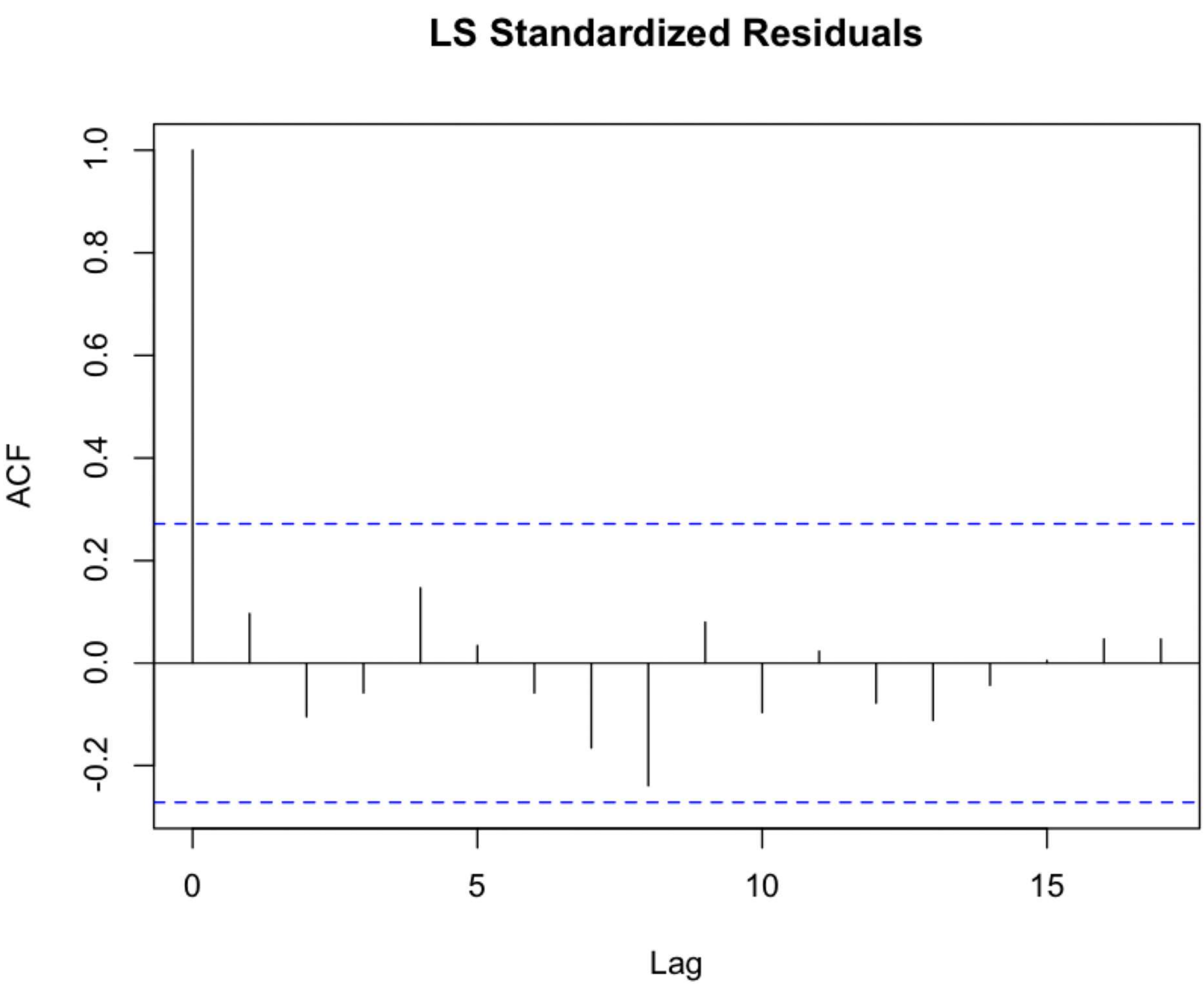
Using LS on the Transformed Data

```
Call:
lm(formula = Ystar ~ Xstar - 1)

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
Xstar(Intercept)   4.675667   0.238370  19.615  < 2e-16 ***
Xstarlog(Price)   -4.327391   0.562556  -7.692 6.44e-10 ***
XstarPromotion     0.584650   0.167111   3.499  0.00102 **
XstarWeek          0.012517   0.004669   2.681  0.01004 *
```

Matches the earlier GLS output

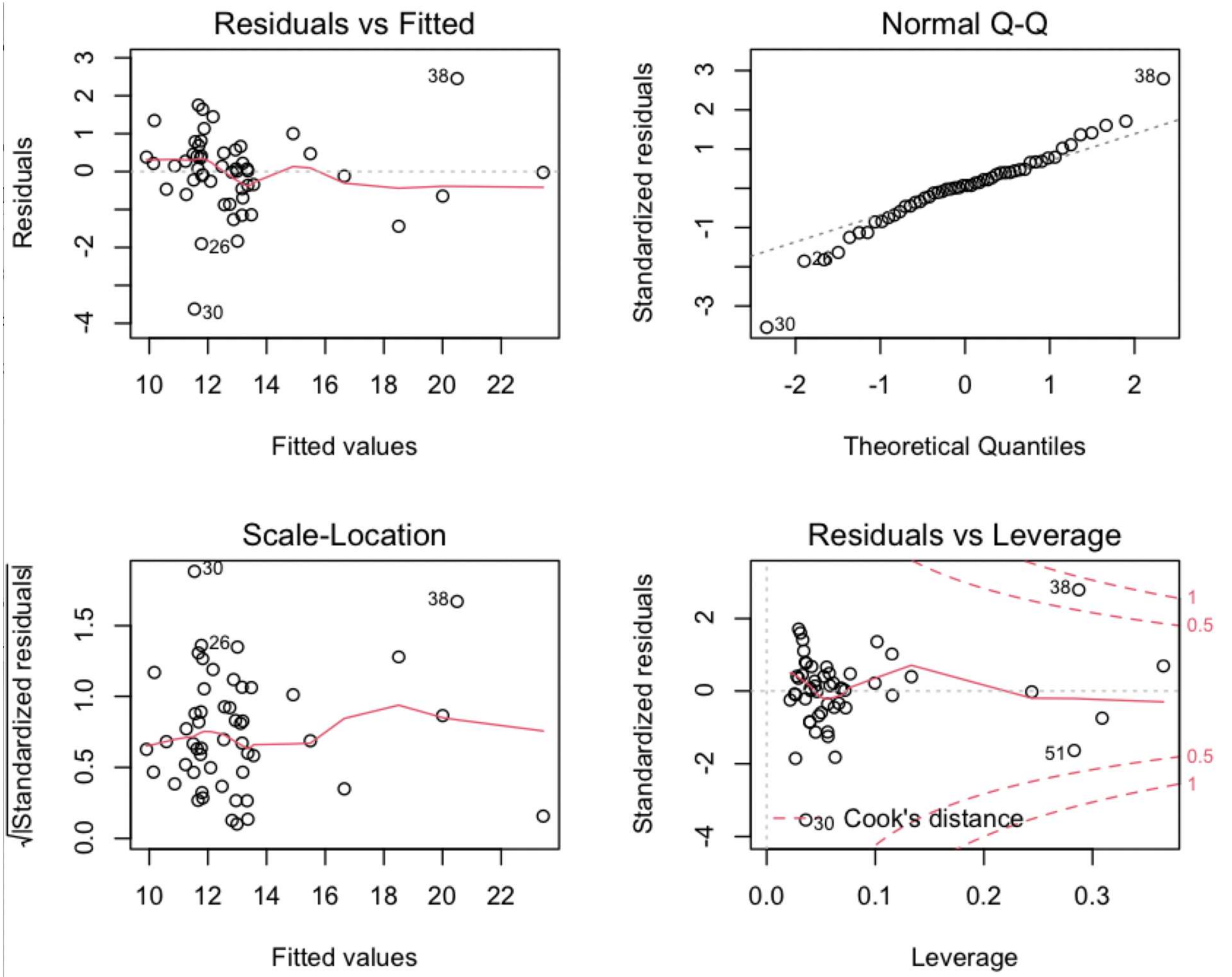


log(Sales)*
Intercept*
log(Price)*
Promotion*
Week*



LS Standardized Residuals
ACF
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
0
5
10
15
Lag

+

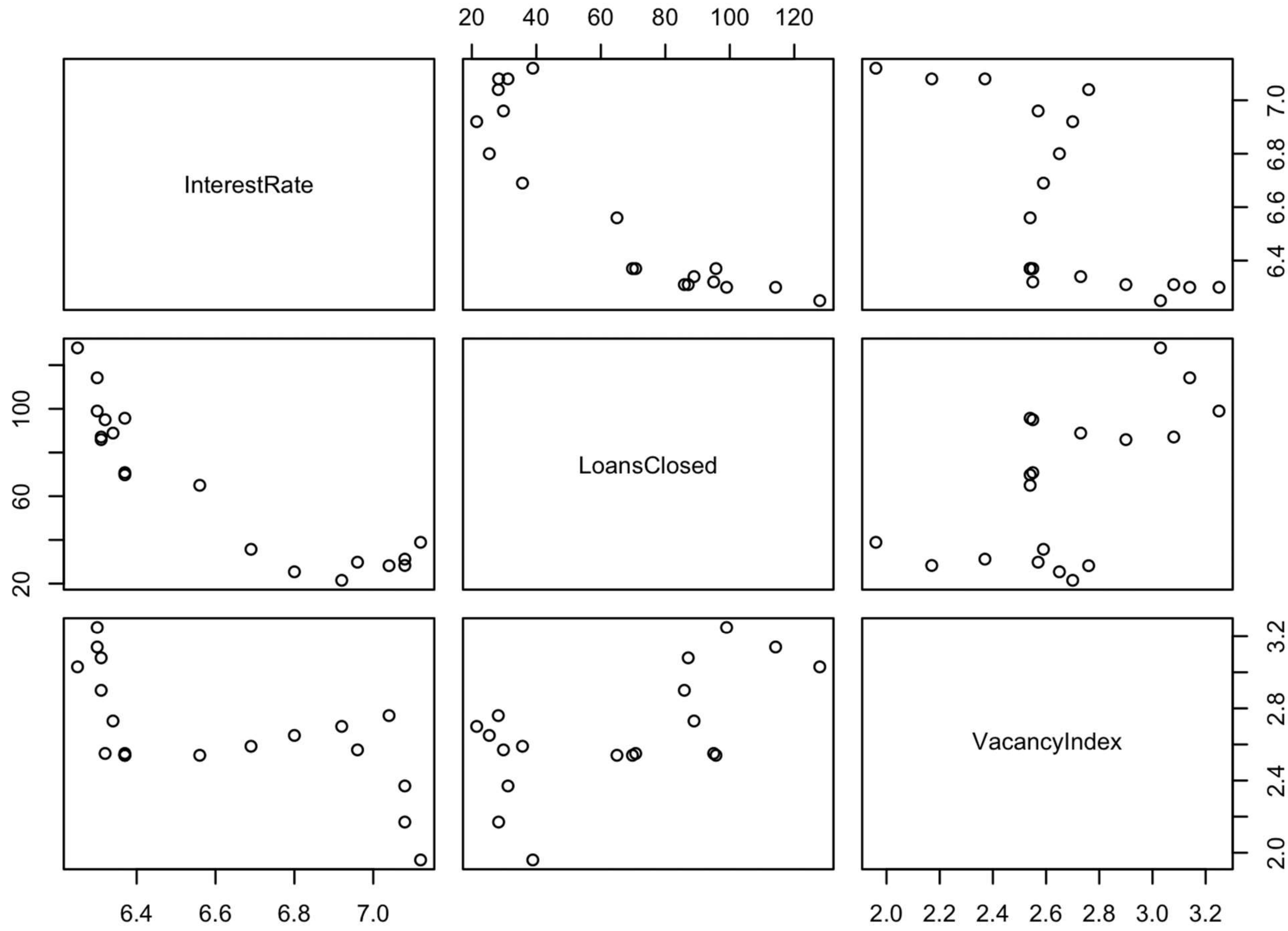

Residuals vs Fitted
Residuals
Fitted values
38
26
30
Normal Q-Q
Standardized residuals
Theoretical Quantiles
38
26
30
Scale-Location
√|Standardized residuals|
Fitted values
30
38
26
Residuals vs Leverage
Standardized residuals
Leverage
38
51
30
Cook's distance
1
0.5
0.5
1

Example

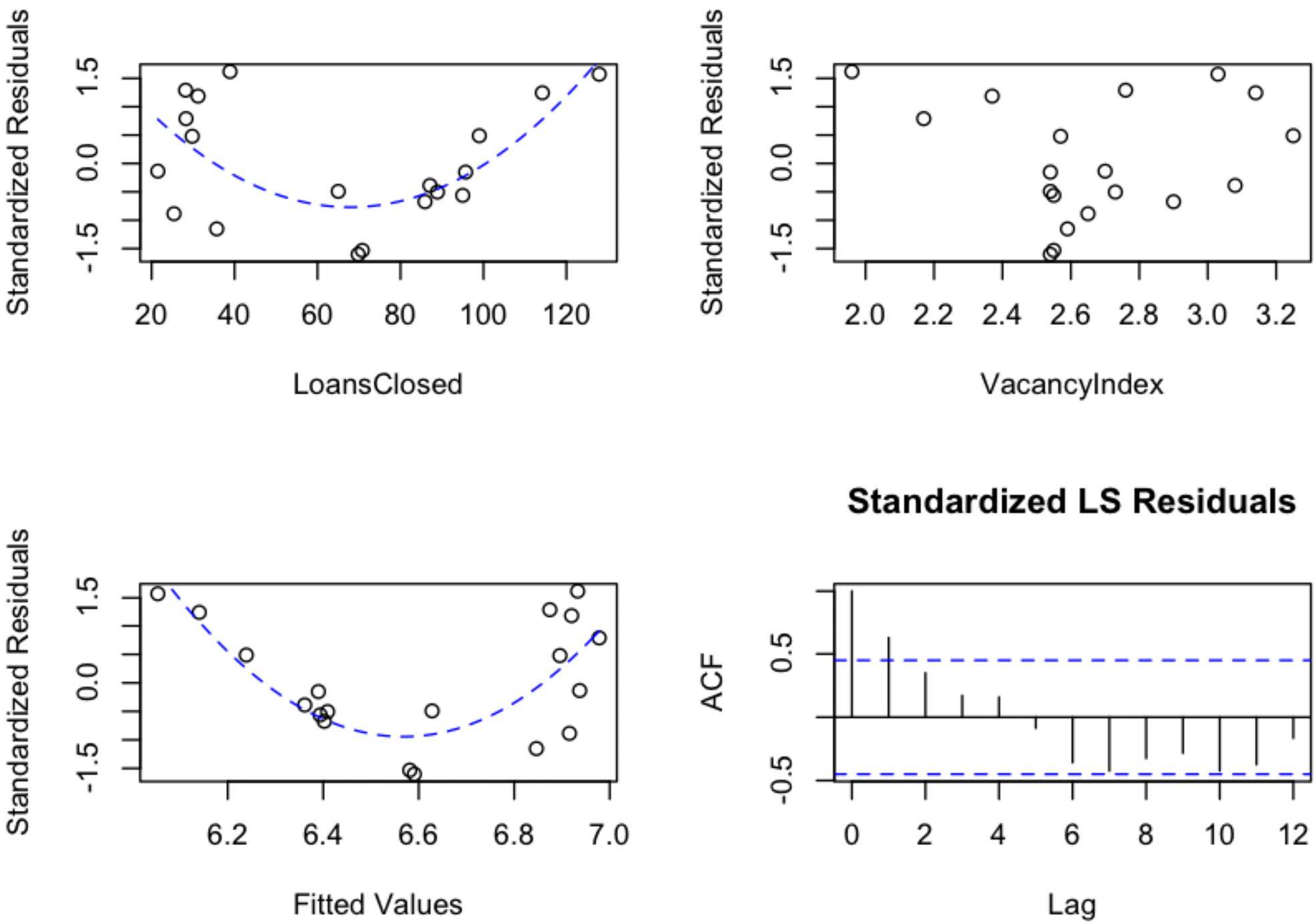
- Loans in the SF Bay Area:
  - Interest rate: Y
  - Amount of loans closed (in millions of dollars): x_1
  - Vacancy index: x_2
  - Goal: Predict interest rate in terms of x_1 and x_2
- We'll start by fitting a naïve model that assumes i.i.d. errors

$$\text{InterestRate}_t = \beta_0 + \beta_1 \text{LoansClosed}_t + \beta_2 \text{VacancyIndex}_t + e$$

- Notice non-linear trends in pairs plot on next slide



InterestRate
LoansClosed
VacancyIndex
20 40 60 80 100 120
6.4 6.6 6.8 7.0
20 60 100
2.0 2.4 2.8 3.2
2.0 2.2 2.4 2.6 2.8 3.0 3.2

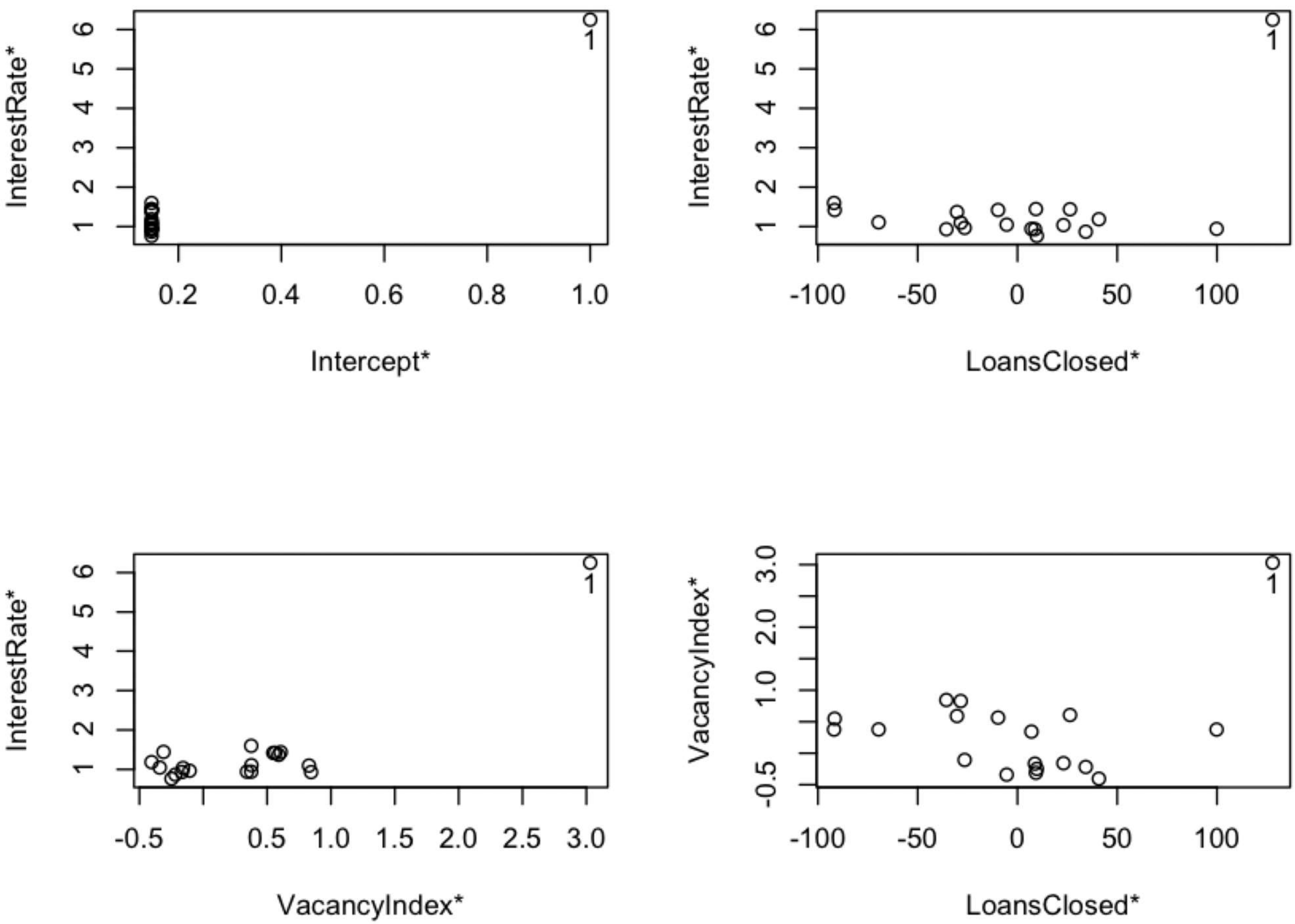

Standardized Residuals
1.5
0.0
-1.5
20
40
60
80
100
120
LoansClosed
Standardized Residuals
1.5
0.0
-1.5
2.0
2.2
2.4
2.6
2.8
3.0
3.2
VacancyIndex
Standardized Residuals
1.5
0.0
-1.5
6.2
6.4
6.6
6.8
7.0
Fitted Values
Standardized LS Residuals
ACF
0.5
-0.5
0
2
4
6
8
10
12
Lag

Now GLS and Autocorrelated Errors

```
Generalized least squares fit by maximum likelihood
  Model: InterestRate ~ LoansClosed + VacancyIndex
  Data: BayArea
        AIC       BIC   logLik
  -35.30833 -30.58613 22.65416

Correlation Structure: AR(1)
 Formula: ~Month
 Parameter estimate(s):
      Phi
0.9572093

Coefficients:
                 Value Std.Error   t-value p-value
(Intercept)   7.122990 0.4182065 17.032232  0.0000
LoansClosed  -0.003432 0.0011940 -2.874452  0.0110
VacancyIndex -0.076340 0.1307842 -0.583710  0.5676
```

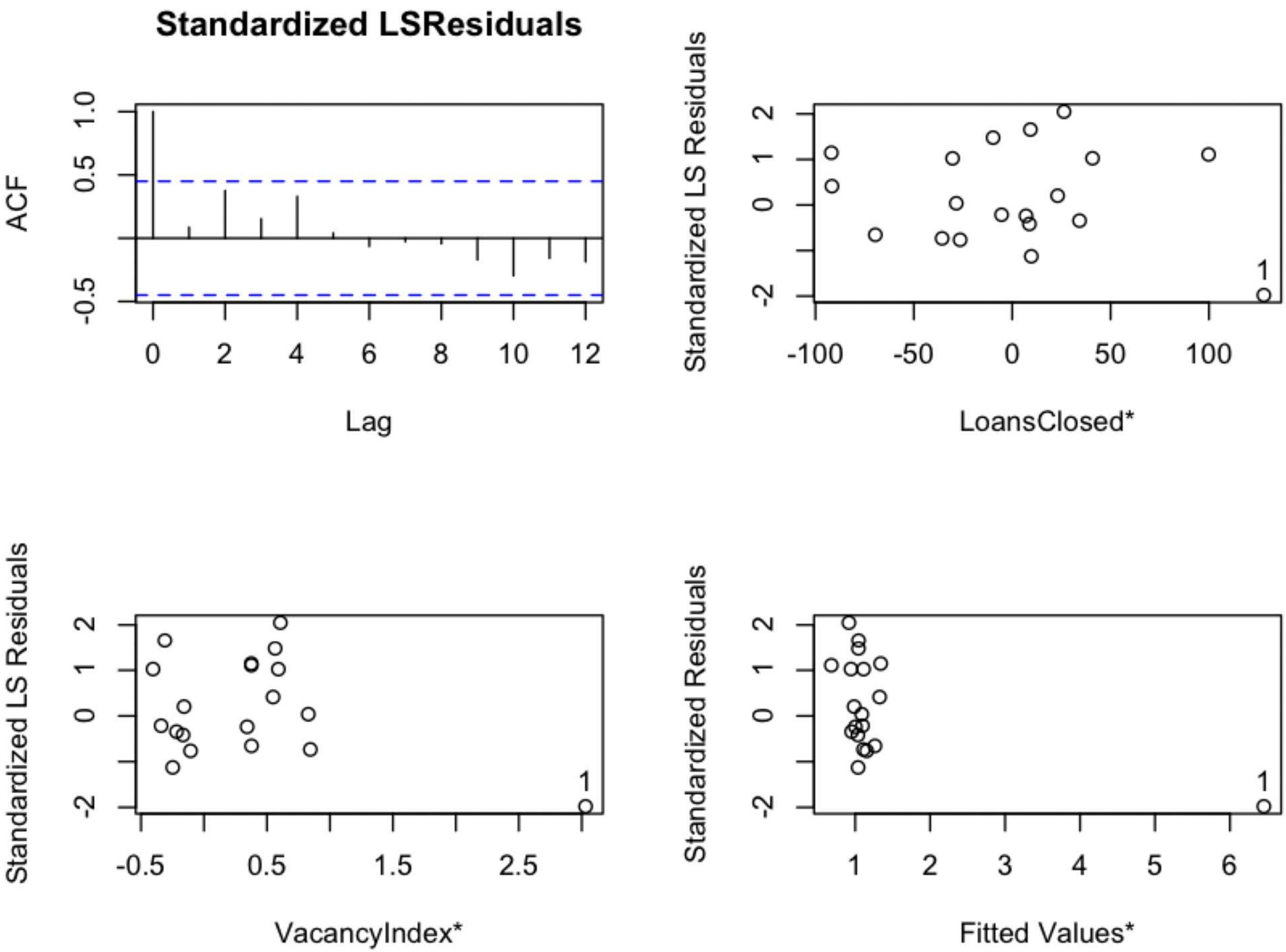
LS on Transformed Variables

```
Call:
lm(formula = ystar ~ xstar - 1)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.27439 -0.12946  0.00934  0.25772  0.52132 

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)    
xstar(Intercept)   7.122990   0.418207  17.032 1.12e-11 ***
xstarLoansClosed  -0.003432   0.001194  -2.874    0.011 *  
xstarVacancyIndex -0.076340   0.130784  -0.584    0.568    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2591 on 16 degrees of freedom
Multiple R-squared:  0.9831,  Adjusted R-squared:  0.9799 
F-statistic: 309.8 on 3 and 16 DF,  p-value: 2.227e-14
```



InterestRate*
Intercept*
LoansClosed*
VacancyIndex*
1

+



Standardized LSResiduals
ACF
Lag
Standardized LS Residuals
LoansClosed*
Standardized LS Residuals
VacancyIndex*
Standardized Residuals
Fitted Values*

Example

- Properly accounting for the correlation fixed the non-linear patterns in the diagnostic plots for the naïve model
- Observation 1 is a highly influential point

@ ≈ 35 min mark (went over Chp 9.5 code)

Finished Wednesday 4/27/22 (Week 14, Lecture 36)