

# Time Series Analysis - 478 - Exam 1

Alex Towell (atowell@siue.edu)

## Contents

<b>Part 1</b>	<b>1</b>
Problem 1.1 . . . . .	1
Part (a) . . . . .	1
Part (b) . . . . .	2
Part (c) . . . . .	2
Problem 1.2 . . . . .	3
Part (a) . . . . .	3
Part (b) . . . . .	3
Part (c) . . . . .	4
Problem 1.3 . . . . .	4
Problem 1.4 . . . . .	6
Part (a) . . . . .	6
Part (b) . . . . .	7
<b>Part 2</b>	<b>9</b>
Problem 2.1 . . . . .	9
Part (a) . . . . .	10
Part (b) . . . . .	10
Part (c) . . . . .	11
Part (d) . . . . .	12
Part (e) . . . . .	12
Problem 2.2 . . . . .	14
Part (a) . . . . .	14
Part (b) . . . . .	15
Part (c) . . . . .	16
Part (d) . . . . .	17
Part (e) . . . . .	17

## Part 1

### Problem 1.1

Suppose that  $\mathbf{Z} := (Z_1, Z_2, Z_3)'$  is a random vector with a mean vector  $\boldsymbol{\mu} = E(\mathbf{Z}) = (0, 1, -1)'$  and a variance-covariance matrix

$$\boldsymbol{\Sigma} = \text{Var}(\mathbf{Z}) = \begin{pmatrix} 1 & -0.5 & 0 \\ -0.5 & 2 & 1.5 \\ 0 & 1.5 & 3 \end{pmatrix}.$$

#### Part (a)

Calculate $E(Z_1 - 3Z_2 - 2Z_3)$ .
------------------------------------

First, given a matrix  $\mathbf{A}$ , we denote the  $(i, j)$ -th element by  $A_{ij}$ . If  $\mathbf{A}$  is a vector, we simplify this notation and denote the  $i$ -th element by  $A_i$ .

The expectation of  $Z_1 - 3Z_2 - 2Z_3$  is given by

$$\begin{aligned} E(Z_1 - 3Z_2 - 2Z_3) &= E(Z_1) - E(3Z_2) - E(2Z_3) \\ &= E(Z_1) - 3E(Z_2) - 2E(Z_3) \\ &= \mu_1 - 3\mu_2 - 2\mu_3. \end{aligned}$$

It is given that  $\boldsymbol{\mu} = (0, 1, -1)$  and thus

$$\begin{aligned} E(Z_1 - 3Z_2 - 2Z_3) &= 0 - 3(1) - 2(-1) \\ &= 0 - 3 + 2 \\ &= -1. \end{aligned}$$

### Part (b)

Calculate  $\text{Var}(2Z_1 + Z_3)$ .

We use the theorems

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2 \text{Cov}(X, Y),$$

and

$$\text{Cov}(aX, bY) = ab \text{Cov}(X, Y).$$

The variance of  $2Z_1 + Z_3$  is given by

$$\begin{aligned} \text{Var}(2Z_1 + Z_3) &= \text{Var}(2Z_1) + \text{Var}(Z_3) + 2 \text{Cov}(2Z_1, Z_3) \\ &= 2^2 \text{Var}(Z_1) + \text{Var}(Z_3) + 2(2 \text{Cov}(Z_1, Z_3)) \\ &= 4\Sigma_{11} + \Sigma_{33} + 4\Sigma_{13} \\ &= 4(1) + (3) + 4(0) \\ &= 7. \end{aligned}$$

Note to Dr. Q: We observe that since  $\Sigma_{13} = 0$ ,  $Z_1$  and  $Z_3$  are linearly uncorrelated. Thus, any function  $g(x, y)$  applied to  $Z_1$  and  $Z_3$  has approximate mean

$$E(g(Z_1, Z_3)) \approx g(\mu_1, \mu_2) + \left. \frac{\partial^2 g}{\partial x^2} \right|_{\mu_1} \Sigma_{11} + \left. \frac{\partial^2 g}{\partial y^2} \right|_{\mu_3} \Sigma_{33}$$

and approximate variance

$$\text{Var}(g(Z_1, Z_3)) \approx \left( \left. \frac{\partial g}{\partial x} \right|_{\mu_1} \right)^2 \Sigma_{11} + \left( \left. \frac{\partial g}{\partial y} \right|_{\mu_3} \right)^2 \Sigma_{33},$$

which are *exact* if  $g$  is a linear function.

### Part (c)

Calculate  $\text{Cov}(3Z_1 - Z_2, Z_2 + 2Z_3)$ .

We use the computational variance theorem  $\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$ , which also means  $E(XY) = \text{Cov}(X, Y) + E(X)E(Y)$ .

The covariance of  $3Z_1 - Z_2$  and  $Z_2 + 2Z_3$  is given by

$$\begin{aligned}\text{Cov}(3Z_1 - Z_2, Z_2 + 2Z_3) &= \text{E}[(3Z_1 - Z_2)(Z_2 + 2Z_3)] - \text{E}(3Z_1 - Z_2) \text{E}(Z_2 + 2Z_3) \\ &= \text{E}(3Z_1Z_2 + 6Z_1Z_3 - Z_2^2 - 2Z_2Z_3) - (3\text{E}(Z_1) - \text{E}(Z_2))(\text{E}(Z_2) + 2\text{E}(Z_3)) \\ &= 3\text{E}(Z_1Z_2) + 6\text{E}(Z_1Z_3) - \text{E}(Z_2^2) - 2\text{E}(Z_2Z_3) - (3\mu_1 - \mu_2)(\mu_2 + 2\mu_3).\end{aligned}$$

Since  $\text{E}(Z_iZ_j) = \Sigma_{ij} + \mu_i\mu_j$ , we may rewrite the above as

$$\begin{aligned}\text{Cov}(3Z_1 - Z_2, Z_2 + 2Z_3) &= 3(\Sigma_{12} + \mu_1\mu_2) + 6(\Sigma_{13} + \mu_1\mu_3) - (\Sigma_{22} + \mu_2^2) - 2(\Sigma_{23} + \mu_2\mu_3) \\ &\quad - (3\mu_1 - \mu_2)(\mu_2 + 2\mu_3).\end{aligned}$$

We are given the values of  $\Sigma_{ij}$  and  $\mu_i$  for  $i, j \in \{1, 2, 3\}$ . We may rewrite the above by making these substitutions,

$$\begin{aligned}\text{Cov}(3Z_1 - Z_2, Z_2 + 2Z_3) &= 3(-0.5 + 0 \cdot 1) + 6(0 + 0 \cdot (-1)) - (2 + 1^2) - \\ &\quad 2(1.5 + 1(-1)) - (3 \cdot 0 - 1)(1 + 2(-1)).\end{aligned}$$

The final calculation results in

$$\text{Cov}(3Z_1 - Z_2, Z_2 + 2Z_3) = -3/2 - 3 - 1 + 3 = -3/2 - 2/2 = -5/2.$$

## Problem 1.2

Let  $\{e_t\}$  be a normal white noise process with mean zero and variance  $\sigma^2$ . Consider the process  $Y_t = e_te_{t-1}$ .

### Part (a)

Calculate  $\text{E}(Y_t)$  and  $\text{Var}(Y_t)$ .

The expectation of  $Y_t$  is given by

$$\text{E}(Y_t) = \text{E}(e_te_{t-1}).$$

By independence,  $\text{E}(Y_t)$  may be rewritten as

$$\text{E}(Y_t) = \text{E}(e_t) \text{E}(e_{t-1}) = 0.$$

The variance of  $Y_t$  is given by

$$\text{Var}(Y_t) = \text{E}(Y_t^2) - \text{E}^2(Y_t).$$

We showed that  $\text{E}(Y_t) = 0$ , thus the variance of  $Y_t$  may be rewritten as

$$\text{Var}(Y_t) = \text{E}((e_te_{t-1})^2) = \text{E}(e_t^2 e_{t-1}^2).$$

Since  $e_t$  and  $e_{t-1}$  are independent,  $e_t^2$  and  $e_{t-1}^2$  are independent, and thus the variance of  $Y_t$  may be rewritten as

$$\text{Var}(Y_t) = \text{E}(e_t^2) \text{E}(e_{t-1}^2) = (\sigma^2 + 0)(\sigma^2 + 0) = \sigma^4.$$

We observe that  $\{Y_t\}$  has constant mean 0 and constant variance  $\sigma^4$ .

### Part (b)

Calculate the ACF (autocorrelation function).

The autocovariance of  $Y_t$  and  $Y_{t-\ell}$ ,  $\ell \geq 0$ , is given by

$$\text{Cov}(Y_t, Y_{t-\ell}) = \text{E}(Y_t Y_{t-\ell}) - \text{E}(Y_t) \text{E}(Y_{t-\ell}).$$

We have already shown in part (a) that  $E(Y_k) = 0$  for any  $k$ , thus

$$\text{Cov}(Y_t, Y_{t-\ell}) = E(Y_t Y_{t-\ell}).$$

Substituting the definition of  $Y_t$  and  $Y_{t-\ell}$  into the above covariance gives

$$\text{Cov}(Y_t, Y_{t-\ell}) = E(e_t e_{t-1} e_{t-\ell} e_{t-\ell-1}).$$

We perform a case analysis to derive the autocovariance for different values of  $\ell$ .

If  $\ell = 0$ , then  $\text{Cov}(Y_t, Y_t) = \text{Var}(Y_t) = \sigma^4$ .

If  $\ell = 1$ , then  $\text{Cov}(Y_t, Y_{t-1}) = E(e_t e_{t-1} e_{t-1} e_{t-2})$ . Let  $W := e_{t-1}^2$ , in which case  $\text{Cov}(Y_t, Y_{t-1}) = E(e_t W e_{t-2})$ . Since these are independent random variables,

$$\text{Cov}(Y_t, Y_{t-1}) = E(e_t) E(W) E(e_{t-2}) = 0 \cdot E(W) \cdot 0 = 0.$$

If  $\ell = 2$ , then

$$\text{Cov}(Y_t, Y_{t-2}) = E(e_t e_{t-1} e_{t-2} e_{t-3}).$$

Since these are independent random variables,

$$\text{Cov}(Y_t, Y_{t-2}) = E(e_t) E(e_{t-1}) E(e_{t-2}) E(e_{t-3}) = 0.$$

Any  $\ell \geq 2$  results in two elements of  $\{Y_t\}$  that have no random noise elements in common and thus also have a covariance of zero.

Since the autocovariance function is symmetric about  $t$ ,  $\text{Cov}(Y_t, Y_{t+\ell}) = \text{Cov}(Y_t, Y_{t-\ell})$ , and thus we see that the autocovariance function is strictly a function of the lag  $\ell$ .

We reparameterize the autocovariance function  $\gamma$  with respect to  $\ell$ ,

$$\gamma_\ell = \begin{cases} \sigma^4 & \ell = 0 \\ 0 & \ell \neq 0. \end{cases}$$

The autocorrelation function  $\text{ls}\rho_\ell$  is defined as  $\gamma_\ell/\gamma_0$ , thus

$$\rho_\ell = \begin{cases} 1 & \ell = 0 \\ 0 & \ell \neq 0. \end{cases}$$

### Part (c)

Is the process weakly stationary? Why?

The process is weakly stationary. It is weakly stationary because its mean is a constant 0, its variance is a constant  $\sigma^4$ , and its autocorrelation function is strictly a function of lag  $\ell$ .

### Problem 1.3

Suppose that we have fit the straight-line regression without intercept  $\hat{y} = \hat{\beta}_1 x_1$ . However, the response  $y$  is in fact affected by a second variable  $x_2$ . So the true regression function is

$$y = \beta_1 x_1 + \beta_2 x_2 + \epsilon.$$

Assume  $\epsilon$ 's are i.i.d. with mean 0 and variance  $\sigma^2$ . Calculate the bias of  $\hat{\beta}_1$  in the original simple linear regression, i.e. calculate  $E(\hat{\beta}_1 - \beta_1)$ .

First, we find an estimator  $\hat{\beta}_1$  given a random sample  $\{Y_i\}$  generated from the model

$$Y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i,$$

except we incorrectly or approximately assume  $Y_i = \beta_1 x_{i1} + \epsilon_i$ . The true statistical error of the  $i$ -th random variable  $Y_i$  is given by

$$\epsilon_i = Y_i - \beta_1 x_{i1} - \beta_2 x_{i2}$$

and we are interested in minimizing the sum of the squares of the statistical errors

$$L = \sum_{i=1}^n \epsilon_i^2.$$

We (incorrectly or approximately) assume  $\epsilon_i = Y_i - \beta_1 x_{i1}$  and parameterize  $L$  with respect to  $\beta_1$ , resulting in the function

$$L(\beta_1) = \sum_{i=1}^n (Y_i - \beta_1 x_{i1})^2.$$

We obtain an estimator for  $\beta_1$  by solving for  $\hat{\beta}_1$  in

$$\left. \frac{\partial L}{\partial \beta_1} \right|_{\hat{\beta}_1} = 0.$$

Thus,

$$\begin{aligned} -2 \sum_i (Y_i - \hat{\beta}_1 x_{i1}) x_{i1} &= 0 \\ \sum_i (Y_i x_{i1} - \hat{\beta}_1 x_{i1}^2) &= 0 \\ \hat{\beta}_1 \sum_i x_{i1}^2 &= \sum_i Y_i x_{i1} \end{aligned}$$

which finally simplifies to

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n Y_i x_{i1}}{\sum_{i=1}^n x_{i1}^2}$$

We are interested in the bias of  $\hat{\beta}_1$ , denoted by

$$b(\hat{\beta}_1) := E(\hat{\beta}_1 - \beta_1).$$

By the linearity of expectation, the bias may be rewritten as

$$b(\hat{\beta}_1) = E(\hat{\beta}_1) - \beta_1.$$

The expectation of  $\hat{\beta}_1$  is given by

$$\begin{aligned} E(\hat{\beta}_1) &= E\left(\frac{\sum_i Y_i x_{i1}}{\sum_i x_{i1}^2}\right) \\ &= \frac{E(\sum_i Y_i x_{i1})}{\sum_i x_{i1}^2} \\ &= \frac{\sum_i E(Y_i x_{i1})}{\sum_i x_{i1}^2} \\ &= \frac{\sum_i x_{i1} E(Y_i)}{\sum_i x_{i1}^2}. \end{aligned}$$

The true model of  $\{Y_i\}$  is given by  $Y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i$ , so

$$\begin{aligned}
E(\hat{\beta}_1) &= \frac{\sum_i x_{i1} E(\beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i)}{\sum_i x_{i1}^2} \\
&= \frac{\sum_i x_{i1} (\beta_1 x_{i1} + \beta_2 x_{i2})}{\sum_i x_{i1}^2} \\
&= \frac{\sum_i \beta_1 x_{i1}^2 + \beta_2 x_{i1} x_{i2}}{\sum_i x_{i1}^2} \\
&= \frac{\sum_i \beta_1 x_{i1}^2}{\sum_i x_{i1}^2} + \frac{\sum_i \beta_2 x_{i1} x_{i2}}{\sum_i x_{i1}^2} \\
&= \beta_1 \frac{\sum_i x_{i1}^2}{\sum_i x_{i1}^2} + \beta_2 \frac{\sum_i x_{i1} x_{i2}}{\sum_i x_{i1}^2} \\
&= \beta_1 + \beta_2 \frac{\sum_i x_{i1} x_{i2}}{\sum_i x_{i1}^2}
\end{aligned}$$

The bias is defined as  $b(\hat{\beta}_1) = E(\hat{\beta}_1) - \beta_1$ , thus

$$b(\hat{\beta}_1) = \beta_2 \frac{\sum_{i=1}^n x_{i1} x_{i2}}{\sum_{i=1}^n x_{i1}^2}.$$

The bias of  $\hat{\beta}_1$  is a function of the observed values  $\{x_{i1}\}$  and  $\{x_{i2}\}$  and the magnitude of  $\beta_2$ .

## Problem 1.4

Consider the simple linear regression model  $y = \beta_0 + \beta_1 x + \epsilon$ , where  $\beta_0$  is known.  $\epsilon$ 's are i.i.d. with mean 0 and variance  $\sigma^2$ .

### Part (a)

Find the least square estimator of  $\beta_1$  in this model.

Since  $\beta_0$  is known, we only need to find an estimator for  $\beta_1$ . The least-squares estimator of  $\beta_1$  is defined as

$$\hat{\beta}_1 = \operatorname{argmin}_{\beta_1} L(\beta_1 | \beta_0)$$

where  $L(\beta_1 | \beta_0) = \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 x_i)^2$ . We obtain the estimator by solving for  $\hat{\beta}_1$  in

$$\left. \frac{\partial L}{\partial \beta_1} \right|_{\hat{\beta}_1} = 0.$$

Thus,

$$\begin{aligned}
-2 \sum_{i=1}^n (Y_i - \beta_0 - \hat{\beta}_1 x_i) x_i &= 0 \\
\sum_{i=1}^n (x_i Y_i - \beta_0 x_i - \hat{\beta}_1 x_i^2) &= 0 \\
\sum_{i=1}^n x_i Y_i - \beta_0 \sum_{i=1}^n x_i - \hat{\beta}_1 \sum_{i=1}^n x_i^2 &= 0 \\
\hat{\beta}_1 \sum_{i=1}^n x_i^2 &= \sum_{i=1}^n x_i Y_i - \beta_0 \sum_{i=1}^n x_i,
\end{aligned}$$

which finally simplifies to

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n x_i Y_i - \beta_0 \sum_{i=1}^n x_i}{\sum_{i=1}^n x_i^2}.$$

Of course, when  $\{Y_i\} = \{y_i\}$ ,  $\hat{\beta}_1$  realizes the particular value

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i - \beta_0 \sum_{i=1}^n x_i}{\sum_{i=1}^n x_i^2}.$$

The expectation of  $\hat{\beta}_1$  is given by

$$\begin{aligned} E(\hat{\beta}_1) &= E \frac{\sum_i x_i Y_i - \beta_0 \sum_i x_i}{\sum_i x_i^2} \\ &= \left( \sum_i x_i^2 \right)^{-1} E \left( \sum_i x_i Y_i - \beta_0 \sum_i x_i \right) \\ &= \left( \sum_i x_i^2 \right)^{-1} \left( \sum_i x_i E(\beta_0 + \beta_1 x_i + \epsilon_i) - \beta_0 \sum_i x_i \right) \\ &= \left( \sum_i x_i^2 \right)^{-1} \left( \beta_0 \sum_i x_i + \beta_1 \sum_i x_i^2 - \beta_0 \sum_i x_i \right) \\ &= \beta_1 \left( \sum_i x_i^2 \right)^{-1} \left( \sum_i x_i^2 \right) \\ &= \beta_1, \end{aligned}$$

which shows that it is unbiased (as expected).

### Part (b)

Construct a  $100(1 - \alpha)\%$  confidence interval for  $\beta_1$ . Compare the interval with the one when  $\beta_0$  is also unknown. Is it narrower?

To construct the confidence interval, we must find the standard deviation of  $\hat{\beta}_1$ . The variance is given by

$$\begin{aligned} \text{Var}(\hat{\beta}_1) &= \text{Var} \left( \frac{\sum_{i=1}^n x_i Y_i - \beta_0 \sum_{i=1}^n x_i}{\sum_{i=1}^n x_i^2} \right) \\ &= \left( \sum_{i=1}^n x_i^2 \right)^{-2} \text{Var} \left( \sum_{i=1}^n x_i Y_i - \beta_0 \sum_{i=1}^n x_i \right) \end{aligned}$$

Since  $\beta_0$  and  $x_i$  for  $i = 1, \dots, n$  are constants, the above simplifies to

$$\begin{aligned}
\text{Var}(\hat{\beta}_1) &= \left( \sum_{i=1}^n x_i^2 \right)^{-2} \text{Var} \left( \sum_{i=1}^n x_i Y_i \right) \\
&= \left( \sum_{i=1}^n x_i^2 \right)^{-2} \sum_{i=1}^n \text{Var}(x_i Y_i) \\
&= \left( \sum_{i=1}^n x_i^2 \right)^{-2} \sum_{i=1}^n x_i^2 \text{Var}(Y_i) \\
&= \left( \sum_{i=1}^n x_i^2 \right)^{-2} \sum_{i=1}^n x_i^2 \text{Var}(\beta_0 + \beta_1 x_i + \epsilon_i) \\
&= \left( \sum_{i=1}^n x_i^2 \right)^{-2} \sum_{i=1}^n x_i^2 \text{Var}(\epsilon_i) \\
&= \left( \sum_{i=1}^n x_i^2 \right)^{-2} \sum_{i=1}^n x_i^2 \sigma^2 \\
&= \sigma^2 \left( \sum_{i=1}^n x_i^2 \right)^{-2} \sum_{i=1}^n x_i^2
\end{aligned}$$

which finally yields the result

$$\text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^n x_i^2}.$$

The standard error is therefore

$$\text{SE}(\hat{\beta}_1) = \frac{\sigma}{\sqrt{\sum_{i=1}^n x_i^2}}.$$

Since  $\hat{\beta}_1$  is a linear combination of standard normal deviates,  $\hat{\beta}_1$  is normally distributed with a mean  $\beta_1$  and a variance  $\text{Var}(\hat{\beta}_1)$ . Thus, a  $100(1 - \alpha)\%$  confidence interval for  $\beta_1$  is

$$\beta_1 \in \left[ \hat{\beta}_1 - z_{\alpha/2} \text{SE}(\hat{\beta}_1), \hat{\beta}_1 + z_{\alpha/2} \text{SE}(\hat{\beta}_1) \right]$$

or, substituting the expression for the standard deviation,

$$\beta_1 \in \left[ \hat{\beta}_1 - \frac{z_{\alpha/2} \sigma}{\sqrt{\sum_i x_i^2}}, \hat{\beta}_1 + \frac{z_{\alpha/2} \sigma}{\sqrt{\sum_i x_i^2}} \right].$$

To compare this estimator with the estimator for unknown  $\beta = (\beta_0, \beta_1)$ , we choose to use the matrix equations for this part as a demonstration of a simpler alternative approach. The unbiased least-squares estimator of  $\beta$  is given by

$$\hat{\beta}_f = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}.$$

where

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$$

and

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}.$$



Performing the matrix calculations, we see that

$$(\mathbf{X}'\mathbf{X}) = \begin{pmatrix} n & \sum_i x_i \\ \sum_i x_i & \sum_i x_i^2 \end{pmatrix}$$

and

$$(\mathbf{X}'\mathbf{X})^{-1} = \frac{1}{n \sum_i x_i^2 - (\sum_i x_i)^2} \begin{pmatrix} \sum_i x_i^2 & -\sum_i x_i \\ -\sum_i x_i & n \end{pmatrix}$$

The variance-covariance matrix of  $\hat{\beta}_f$  is therefore

$$\Sigma = \text{Var}(\hat{\beta}_f) = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}$$

and therefore the variance of  $\hat{\beta}_{1,f}$  is given by

$$\begin{aligned} \text{Var}(\hat{\beta}_{1,f}) &= \sigma^2 \Sigma_{22} \\ &= \frac{n\sigma^2}{n \sum_i x_i^2 - (\sum_i x_i)^2} \\ &= \frac{\sigma^2}{\sum_i x_i^2 - \frac{1}{n} (\sum_i x_i)^2} \end{aligned}$$

Recall that  $\text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum_i x_i^2}$ . The ratio of  $\text{Var}(\hat{\beta}_1)$  to  $\text{Var}(\hat{\beta}_{1,f})$  is therefore

$$\begin{aligned} w &= \frac{\sum_i x_i^2 - \frac{1}{n} (\sum_i x_i)^2}{\sum_i x_i^2} \\ &= 1 - \frac{(\sum_i x_i)^2}{\sum_i x_i^2} \\ &= 1 - \frac{n\bar{x}^2}{\sum_i x_i^2}, \end{aligned}$$

which implies  $0 < w < 1$ , i.e.,  $\text{Var}(\hat{\beta}_{1,f})$  is larger than  $\text{Var}(\hat{\beta}_1)$ . In particular, since  $w = \frac{\text{Var}(\hat{\beta}_1)}{\text{Var}(\hat{\beta}_{1,f})}$ ,

$$\sigma_{\hat{\beta}_1} = \sqrt{w} \sigma_{\hat{\beta}_{1,f}}$$

and thus we may conclude that the confidence interval for  $\hat{\beta}_1$  is a factor  $\sqrt{w}$  as wide as the confidence interval for  $\hat{\beta}_{1,f}$ , where  $\sqrt{w} < 1$ .

It was immediately obvious that knowing the value of  $\beta_0$  should reduce the uncertainty of  $\beta_1$  given a sample, but I thought this proof was fairly interesting.

Finally, to answer the question, since a larger variance yields a larger confidence interval, the confidence interval for  $\hat{\beta}_1$  is narrower than the confidence interval for  $\hat{\beta}_{1,f}$ .

## Part 2

### Problem 2.1

The `EmployeeData` data set gives the number of employees (in thousands) for a metal fabricator and one of their primary vendors for each month over a 5-year period. You may find the data in .txt file on blackboard and read the data into R using `read.table` command.

### Part (a)

Fit a simple linear model to the data, where  $y_t$  is the number of employees during time period  $t$  at the metal fabricator and  $x_t$  is the number of employees at the vendor. Report the ANOVA table and summary for the model coefficients.

```
emp_data <- read.table("EmployeeData.txt", header=TRUE)

# fit a multiple regression model
ols.fit <- lm(metal~vendor, data=emp_data)

# get details from the regression output
summary(ols.fit)

##
## Call:
## lm(formula = metal ~ vendor, data = emp_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.2348 -1.2393 -0.0311  1.0022  3.7077
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.847911    3.299962   0.863   0.392
## vendor       0.122442    0.009423  12.994 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.59 on 58 degrees of freedom
## Multiple R-squared:  0.7443, Adjusted R-squared:  0.7399
## F-statistic: 168.8 on 1 and 58 DF,  p-value: < 2.2e-16

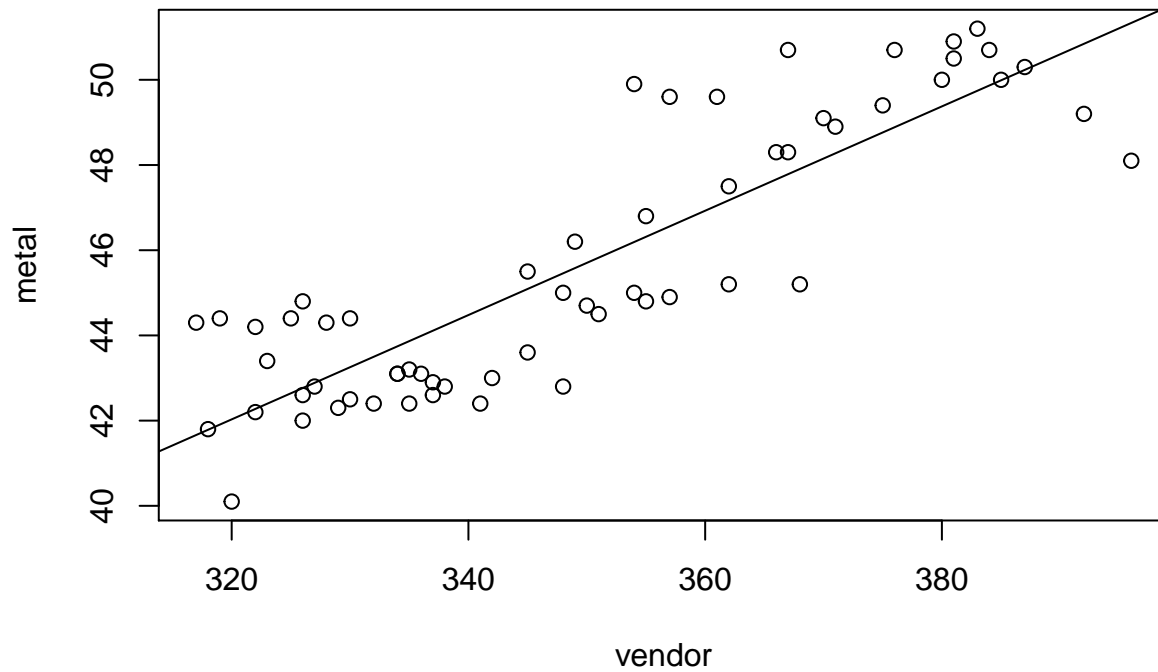
# get the anova table
anova(ols.fit)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
vendor	1	426.7163	426.716336	168.834	0
Residuals	58	146.5910	2.527431	NA	NA

### Part (b)

Plot of the number of employees at the fabricator versus the number of employees at the vendor with the ordinary least squares regression line overlaid.

```
# vendor and metal seem to be positively correlated.
with(emp_data,plot(vendor,metal))
abline(ols.fit)
```

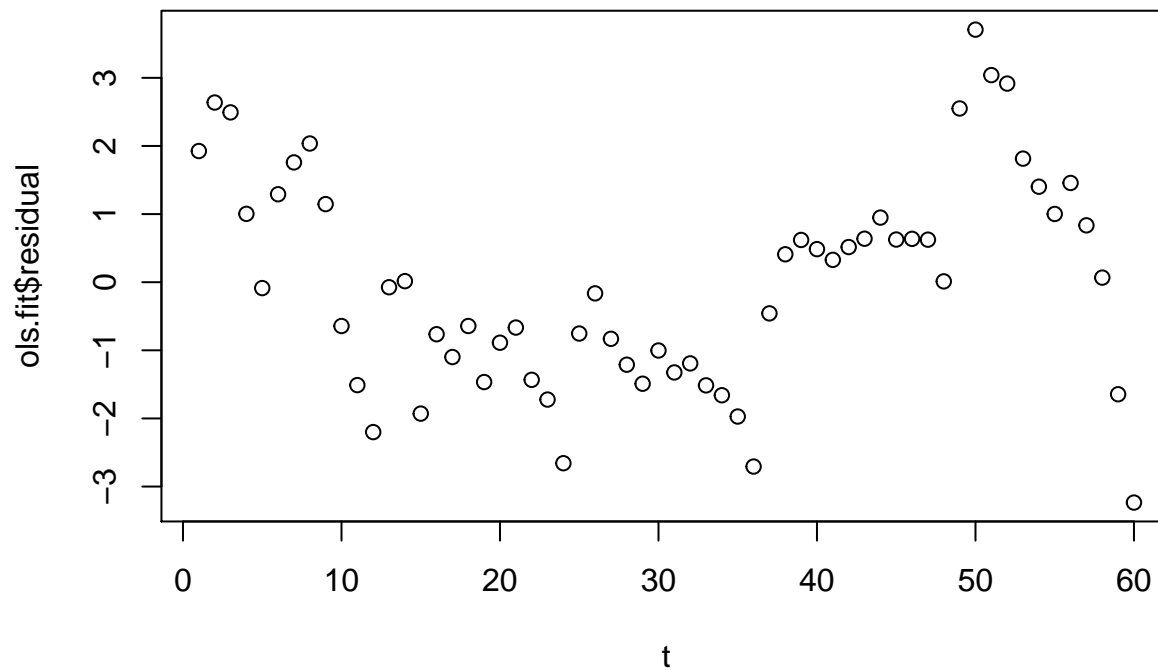


### Part (c)

Plot of the residuals versus  $t$  (the time ordering). Does it look random?

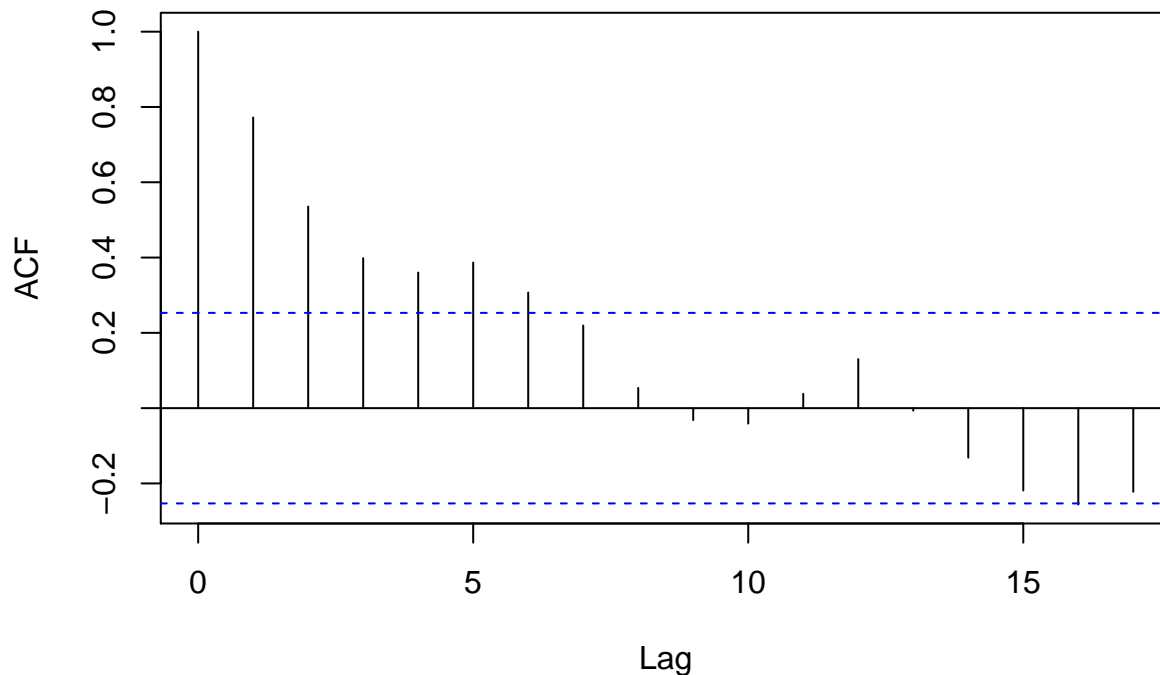
Here's the time-ordering plot of the residuals.

```
N=nrow(emp_data)
t=1:N
plot(t,ols.fit$residual)
```



```
acf(ols.fit$residual)
```

### Series `ols.fit$residual`



The residuals do not appear to be i.i.d. normally distributed around 0.

#### Part (d)

Conduct a Durbin-Watson test to determine the correlation in the residuals. Comment on your conclusion.

```
library("lmtest")
dwtest(ols.fit)
```

```
##
## Durbin-Watson test
##
## data:  ols.fit
## DW = 0.35924, p-value < 2.2e-16
## alternative hypothesis: true autocorrelation is greater than 0
```

If the test statistic DW is around 2, there is strong evidence that there is no autocorrelation. In this case, the statistic is quite small, and so we have strong evidence to reject the null hypothesis of no autocorrelation.

#### Part (e)

Use one iteration of the Cochrane-Orcutt procedure to estimate the regression coefficients. Also calculate the standard errors of the coefficients. Are the standard errors (from the Cochrane-Orcutt procedure) larger than the ones from simple linear regression?

```
# calculate phi for the Cochrane Method
phi.hat=lm(ols.fit$residual[2:N]~0+ols.fit$residual[1:N-1])$coeff

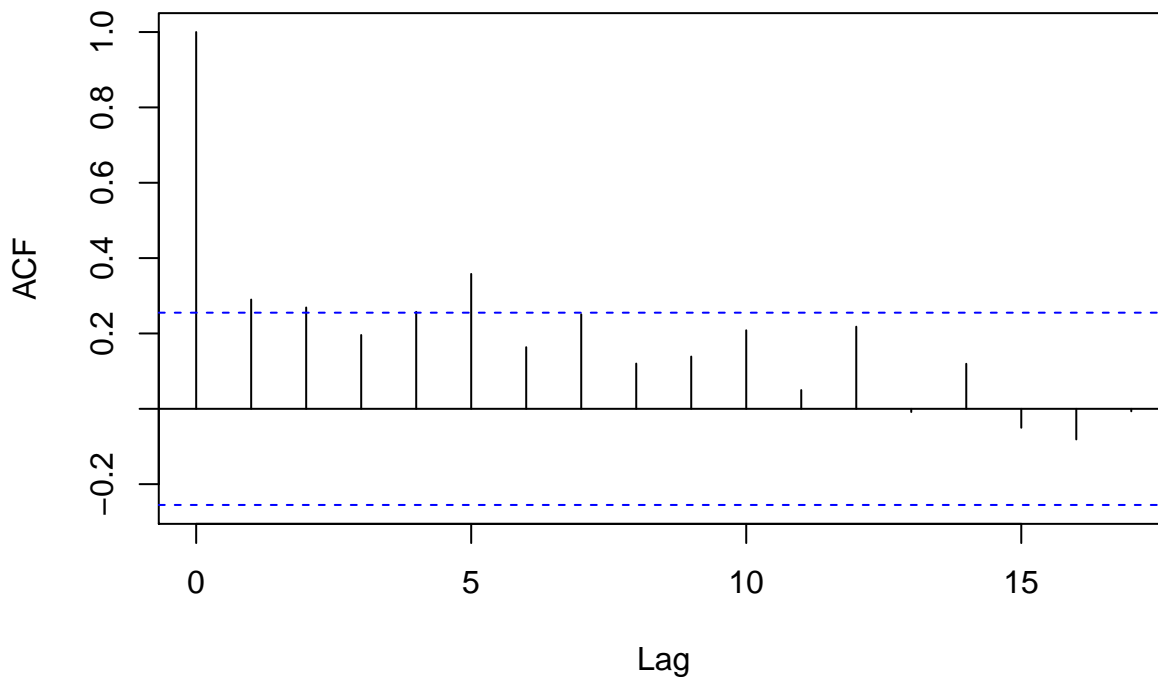
# transform y and x according to the Cochrane Method
y.trans=emp_data$metal[2:N]-phi.hat*emp_data$metal[1:N-1]
```

```
x.trans=emp_data$vendor[2:N]-phi.hat*emp_data$vendor[1:N-1]

# fit OLS regression with transformed data
coch.or=lm(y.trans~x.trans)
summary(coch.or)

##
## Call:
## lm(formula = y.trans ~ x.trans)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.1944 -0.4425  0.1461  0.5125  1.2218
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.87560    0.78655   6.199 6.78e-08 ***
## x.trans       0.04795    0.01300   3.688 0.000505 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7342 on 57 degrees of freedom
## Multiple R-squared:  0.1927, Adjusted R-squared:  0.1785
## F-statistic: 13.6 on 1 and 57 DF, p-value: 0.0005054
acf(coch.or$residual)
```

### Series coch.or\$residual



The standard error of the simple linear regression model for the intercept was 3.299962. The standard error for the intercept in the Cochrane-Orcutt model is significantly smaller at 0.78655.

The standard error of the simpler linear regression model for the slope is 0.009423. The standard error for the intercept in the Cochrane-Orcutt model is slightly larger at 0.01300.

## Problem 2.2

The following analysis are based on the data in `HomePrice.txt` file on blackboard. You may read the data into R using `read.table` command. This `HomePrice` dataset has the following variables:

- $Y$  = sale price of home
- $X_1$  = logged square footage of home
- $X_2$  = logged square footage of the lot

### Part (a)

Fit an ordinary linear regression model,  $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$ . Report the ANOVA table and summary for the model coefficients.

```
hp_data <- read.table("HomePrice.txt", header=TRUE)
colnames(hp_data) = c("t", "Y", "X1", "X2")

# fit a multiple regression model
hp_model <- lm(Y~X1+X2, data=hp_data)

# get details from the regression output
summary(hp_model)

##
## Call:
## lm(formula = Y ~ X1 + X2, data = hp_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -228421  -38178   -5506   25494  383423
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.027e+05  1.265e+04  -8.121 3.39e-15 ***
## X1           1.560e+02  4.871e+00  32.019 < 2e-16 ***
## X2           1.151e+00  2.964e-01   3.882 0.000117 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 78070 on 519 degrees of freedom
## Multiple R-squared:  0.6808, Adjusted R-squared:  0.6796
## F-statistic: 553.5 on 2 and 519 DF, p-value: < 2.2e-16

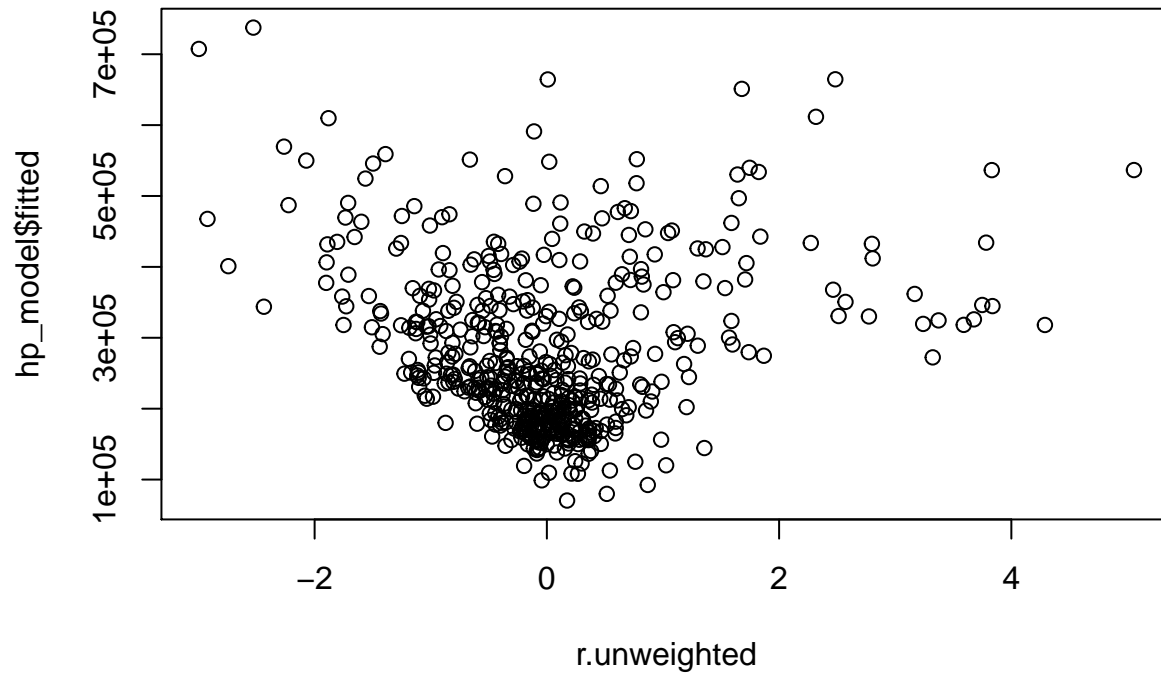
# get the anova table
anova(hp_model)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
X1	1	6.655486e+12	6.655486e+12	1091.87527	0.0000000
X2	1	9.187986e+10	9.187986e+10	15.07348	0.0001168
Residuals	519	3.163546e+12	6.095464e+09	NA	NA

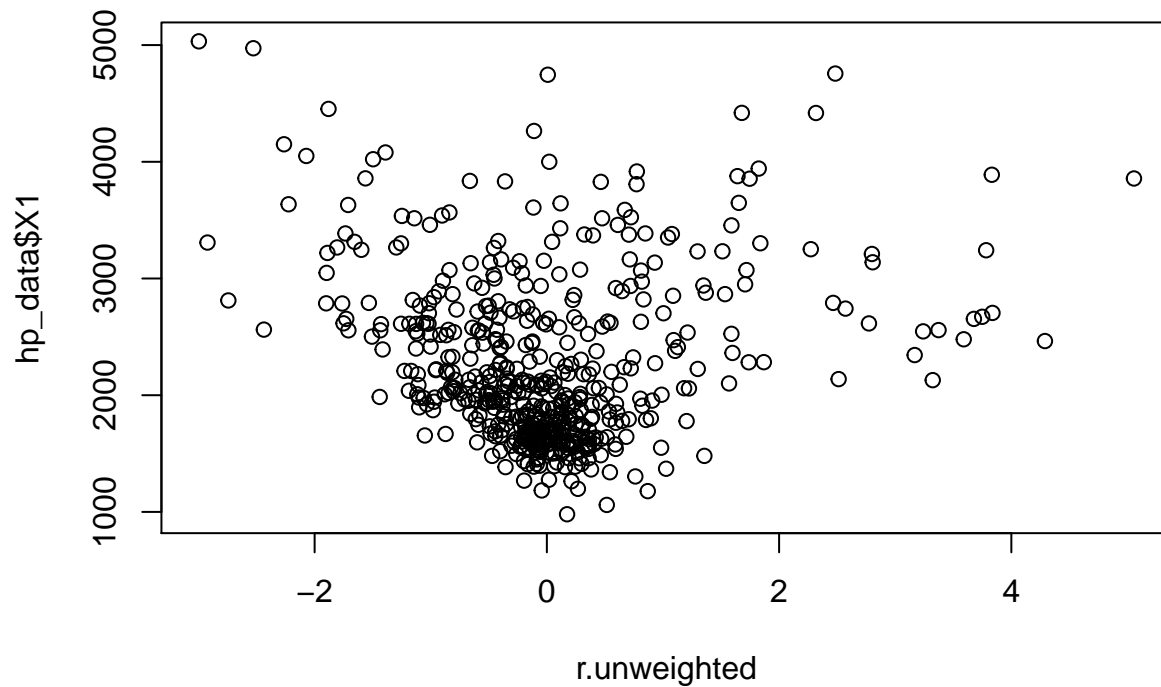
Part (b)

Plot of the OLS residuals versus OLS fitted values. Comment on any pattern you see.

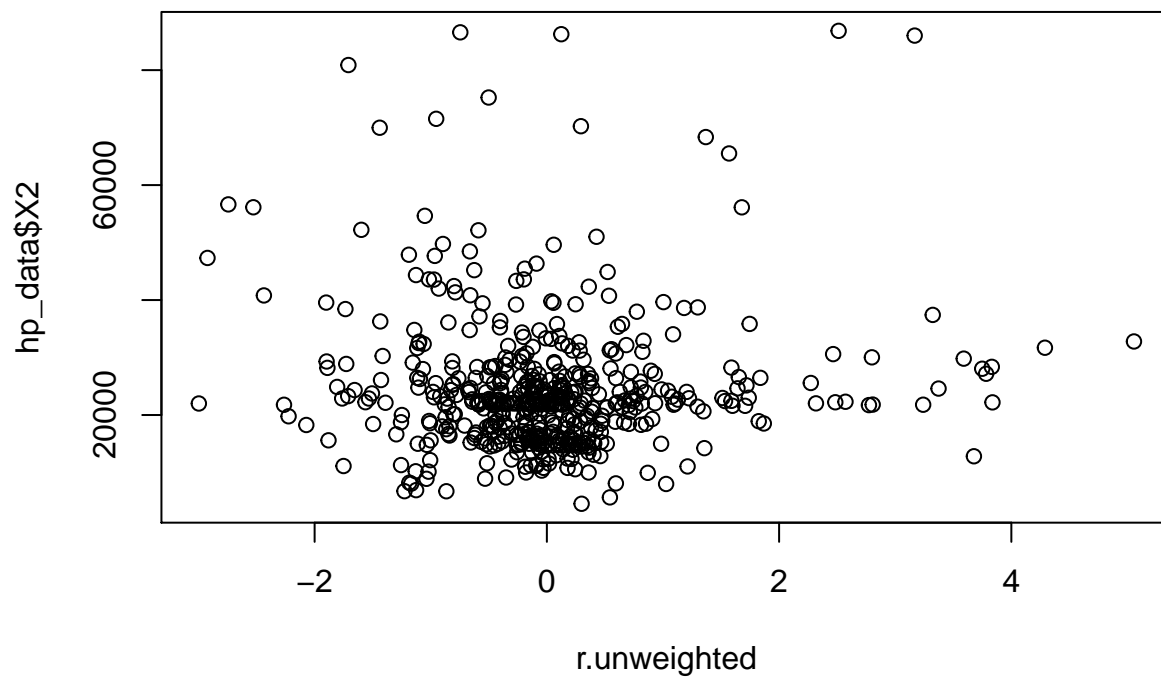
```
# studentized residuals  
r.unweighted = rstudent(hp_model)  
  
# the following plot of the fitted values vs the residuals suggests  
# non-constant variance. the variance seems to be increasing with respect  
# to y. the residuals do seem to have a zero mean though.  
plot(r.unweighted, hp_model$fitted)
```



```
# variance seems more constant with respect to home price (X1)  
plot(r.unweighted, hp_data$X1)
```



```
plot(r.unweighted, hp_data$X2)
```



The studentized residuals of the regression seems to have a non-constant variance with respect to the fitted values  $\hat{y}$ ,  $X_1$ , and  $X_2$ . Primarily, they exhibit the fanning out characteristic that you typically see with non-constant variance.

### Part (c)

Calculate the absolute values of the OLS residuals. Regress the absolute values of the OLS residuals versus the OLS fitted values and store the fitted values from this regression.



```
# the absolute value of the OLS residuals.
abs_residuals = abs(residuals(hp_model))

# we're fitting
#  $s(i) = \gamma_0 + \gamma_1 y(i) + \zeta(i)$ 
# where  $\zeta(i)$  is the random error.
abs_residuals_fit=lm(abs_residuals~hp_model$fitted)
```

#### Part (d)

Calculate weights equal to  $1/\hat{\epsilon}^2$ , where  $\hat{\epsilon}$  are the fitted values from the regression in the last step. Using these weights this time in a weighted least squares regression. Report the ANOVA table and summary for the model coefficients.

```
# weighted least squares. we're taking the squared reciprocal of the estimated
# residuals from the regression model as the weight matrix.
wts=1/(fitted(abs_residuals_fit))^2

# fit the weighted regression model to the data
hp_model.weighted=lm(Y~X1+X2, data=hp_data,weights=wts)

anova(hp_model.weighted)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
X1	1	4095.7933798	4095.7933798	1754.0375863	0.0000000
X2	1	0.7163133	0.7163133	0.3067637	0.5799115
Residuals	519	1211.8992094	2.3350659	NA	NA

```
summary(hp_model.weighted)
```

```
##
## Call:
## lm(formula = Y ~ X1 + X2, data = hp_data, weights = wts)
##
## Weighted Residuals:
##      Min       1Q   Median       3Q      Max
## -9.4644 -0.9364 -0.2118  0.6141  8.0706
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -8918.7876   3619.3749  -2.464   0.0141 *
## X1           123.1438     3.3186   37.107  <2e-16 ***
## X2           -0.1274     0.2300   -0.554   0.5799
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.528 on 519 degrees of freedom
## Multiple R-squared:  0.7717, Adjusted R-squared:  0.7708
## F-statistic: 877.2 on 2 and 519 DF, p-value: < 2.2e-16
```

#### Part (e)

Plot of the WLS residuals versus WLS fitted values. Does it look random now?

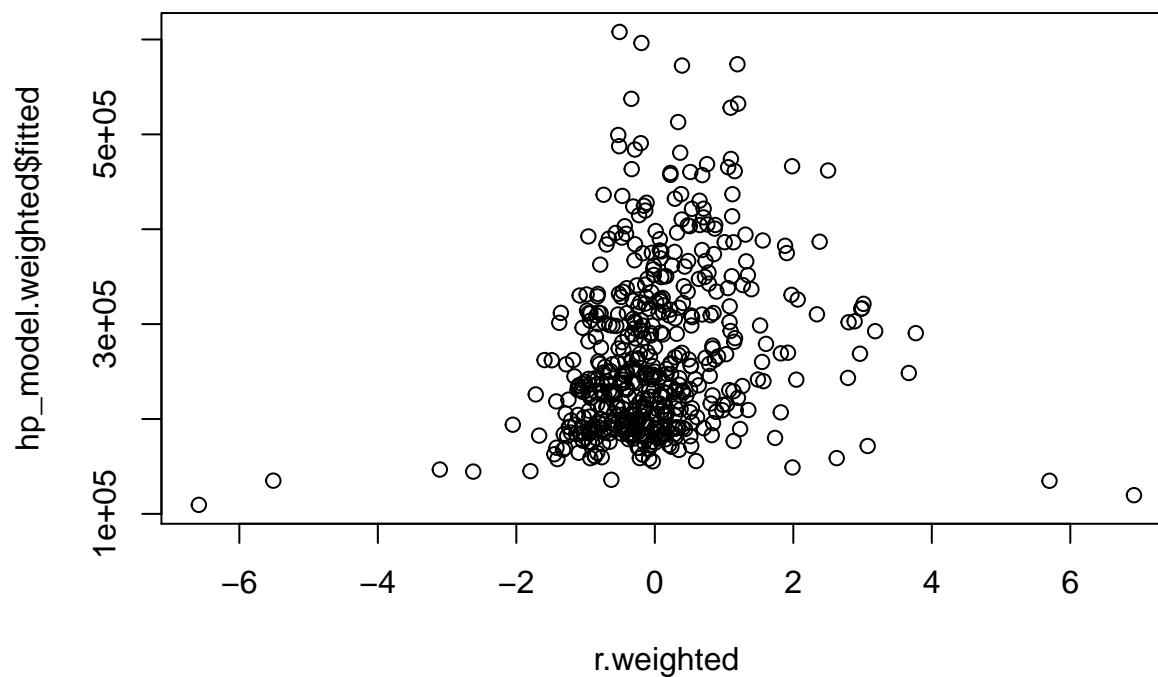
```
# weighted residual analysis
```

```
# studentized residuals
```

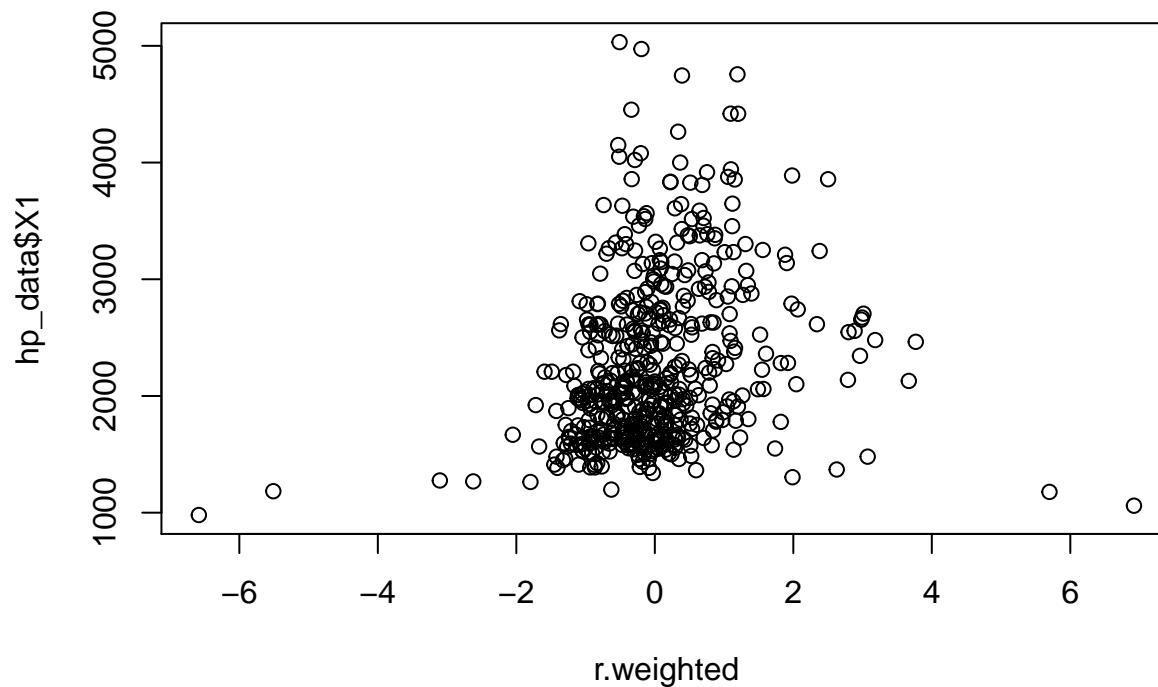
```
r.weighted = rstudent(hp_model.weighted)
```

```
# plot studentized residual vs fitted values
```

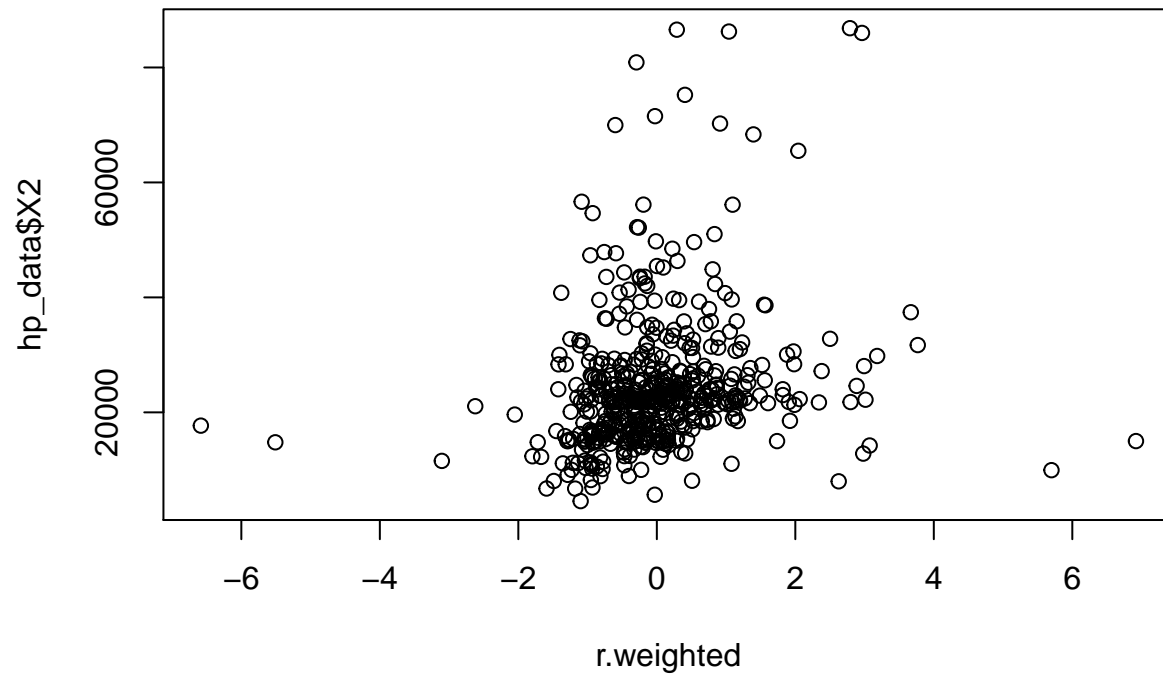
```
plot(r.weighted, hp_model.weighted$fitted)
```



```
plot(r.weighted, hp_data$X1)
```



```
plot(r.weighted, hp_data$X2)
```



The studentized residuals of the weighted regression seems to have a more constant variance with respect to the fitted values  $\hat{y}$ ,  $X_1$ , and  $X_2$ . Primarily, they do not exhibit the fanning out characteristic that you typically see with non-constant variance.

In conclusion, the weighted regression seems to generate residuals that are more presentative of random white noise with zero mean and constant variance.